UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

Mestrado em Econometria Aplicada e Previsão

Testes de especificação ao modelo de *Hazard* Proporcional

Daniel Alexandre Baptista Dias

Orientação: Doutor José Manuel de Matos Passos

Júri:
Presidente:
Doutor Artur Carlos Barros da Silva Lopes, professor associado do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.

Vogais:
Doutor José António Ferreira Machado, professor catedrático da Faculdade de Economia da Universidade Nova de Lisboa.

Doutor José Manuel de Matos Passos, professor auxiliar do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.

Doutor Montezuma Boaventura Guimarães Dumangane, professor auxiliar do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.

Fevereiro/2005

UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

Mestrado em Econometria Aplicada e Previsão

Testes de especificação ao modelo de *Hazard* Proporcional

Daniel Alexandre Baptista Dias

Orientação: Doutor José Manuel de Matos Passos

Júri:
Presidente:
Doutor Artur Carlos Barros da Silva Lopes, professor associado do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.

Vogais:
Doutor José António Ferreira Machado, professor catedrático da Faculdade de Economia da Universidade Nova de Lisboa.

Doutor José Manuel de Matos Passos, professor auxiliar do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.

Doutor Montezuma Boaventura Guimarães Dumangane, professor auxiliar do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.

Fevereiro/2005

Resumo

O estudo dos factores que influenciam o tempo entre dois acontecimentos é um tópico comum a vários ramos das ciências económicas. Na literatura, este tipo de análise é conhecida por *"Análise da Duração"* e os modelos estatísticos/econométricos que lhe estão associados são os *"Modelos de Duração"*. No trabalho que se segue iremos apresentar um conjunto de testes de especificação ao modelo de *Hazard* Proporcional assim como iremos analisar as propriedades de amostras finitas de testes de especificação gerais, em particular, testes de momentos condicionais.

Palavras chave: Modelos de Duração, *Hazard* Proporcional, Testes de Especificação, Modelo de Cox, Momentos Condicionais, Simulação Monte Carlo.

Classificação JEL: C12, C14, C15, C41, C52.

Abstract

The analysis of factors influencing the time between events is a topic common to several areas of the economic sciences. In the literature, this type of analysis is known as Duration Analysis and the inherent econometric models are the Duration Models. In this work, we will present some specification tests of the Proportional Hazard model and we will also analyse the finite sample properties of general specification tests, in particular, conditional moments tests.

Keywords: Duration Models, Proportional Hazard, Specification Tests, Cox Model, Conditional Moments, Monte Carlo simulation.

JEL Classification: C12, C14, C15, C41, C52.

Agradecimentos

Em primeiro lugar, desejo agradecer ao Professor Doutor José Passos, orientador desta dissertação, pelos preciosos comentários e sugestões que efectuou ao longo da elaboração do presente trabalho e pela forma dedicada como acompanhou a sua realização.

Quero agradecer ao Professor Doutor João Santos Silva pelos vários conselhos e esclarecimentos que me deu ao longo da elaboração deste trabalho.

Agradeço também aos meus pais por todo o apoio que me deram desde o ínicio deste meu projecto pessoal.

Uma referência ainda para o Banco de Portugal, em particular para o Departamento de Estudos Económicos, por todo o apoio que me prestou em termos de meios informáticos.

Índice

Lista de Quadros

1 Introdução

O estudo dos factores que influenciam o tempo entre dois acontecimentos é um tópico comum a vários ramos das ciências económicas, por exemplo, saber se a escolaridade de um indivíduo influencia a duração do desemprego, ou, saber se a estrutura de capital de uma empresa influencia a sua longevidade. Na literatura, este tipo de análise é conhecida por *"Análise da Duração"* e os modelos estatísticos/econométricos que lhe estão associados são os *"Modelos de Duração"*. Como principais características dos dados de duração salientamos o facto de (i) serem sempre positivos e (ii) poderem ser censurados.

Tal como em qualquer outro tipo de modelação econométrica, a modelação da duração consiste em encontrar relações entre uma variável dependente, neste caso o tempo decorrido ou duração, e um conjunto de variáveis explicativas ou regressores. O método natural de o fazer seria através de OLS mas, caso a variável dependente seja censurada, as hipóteses deste estimador são violadas. Esta situação faz com que não seja possível estimar consistentemente os parâmetros relevantes. Dada esta restrição, existem dois tipos de modelação alternativos. O primeiro consiste em especificar a distribuição condicional ou, de forma equivalente, a função *hazard* condicional. Apesar da equivalência, em economia é mais usual especificar a função *hazard* condicional do que a distribuição condicional. A especificação da função *hazard* pode ser paramétrica (casos da distribuição weibull, gama, log-logistica, etc) ou semi-paramétrica (caso do modelo de Cox). Na modelação pela *hazard* é ainda comum classificar os modelos em modelos PH (*Proportional Hazard*) e em modelos AFT (*Accelerated Failure Time*). A segunda opção de modelação consiste em especificar momentos condicionais da distribuição, por exemplo, a média ou a mediana condicional.

A grande vantagem da modelação paramétrica é ser possível obter um conjunto de resultados que lhe são inerentes, nomeadamente, o valor esperado condicional, percentis condicionais, probabilidades condicionais, eficiência máxima na estimação, etc. Em termos de desvantagens, pode-se dizer que existe apenas uma, ou seja, raras são as vezes em que se conhece a distribuição condicional da duração. Nestes casos, a inferência que se pretenda fazer poderá estar errada, pelo que, testar a adequabilidade do modelo condicional é fundamental.

No que diz respeito à estimação de momentos condicionais[1], as principais vantagens são a robustez destes estimadores à existência de censura e a alguns tipos de má especificação, por exemplo, a certas formas de heterogeneidade negligenciada, e, ao mesmo tempo, permitirem a estimação de parâmetros diferentes para cada momento condicional. As principais desvantagens são a impossiblidade de calcular probabilidades e, em alguns casos, não permitirem a estimação do valor esperado condicional[2].

Apesar da diversidade de técnicas e métodos de modelação da duração, o presente trabalho apenas incidirá sobre um modelo em particular, o modelo de *Hazard* Proporcional. Em relação a este modelo procurou-se atingir dois objectivos: (i) apresentar testes de especificação relativos às hipótese do modelo de *Hazard* Proporcional e (ii) através de simulação de Monte Carlo, analisar as propriedades em

[1] As principais referências neste tipo de métodos são Koenker e Basset (1978) e Powell (1984 e 1986).

[2] Existem estimadores que permitem a estimação do valor esperado condicional sem especificar qualquer distribuição condicional, um destes casos é o estimador proposto por Buckley e James (1979).

amostras finitas dos testes de momentos condicionais propostos por Tauchen (1985) e Newey (1985) e por Horowitz e Neumann (1992).

O texto que se segue está organizado da seguinte forma: a secção 2 apresenta o modelo de *Hazard* Proporcional e faz uma breve introdução aos principais conceitos relacionados com modelos de duração que irão ser necessários ao longo do texto. A secção 3 apresenta os testes de especificação ao modelo de *Hazard* Proporcional. A secção 4 apresenta os resultados, obtidos através de simulação de Monte Carlo, do estudo das propriedades em pequenas amostras de testes de momentos condicionais. Finalmente, a secção 5 conclui.

2 Modelo de *Hazard* Proporcional

2.1 Conceitos fundamentais sobre Modelos de Duração

Seja T uma variável aleatória não negativa, X um vector de variáveis aleatórias observadas[3] de dimensão $k \times 1$, $H(x)$ o vector das funções de distribuição de probabilidade dos regressores contidos em X, V um escalar aleatório não observado referente à heterogeneidade negligenciada, $G(v)$ a função de distribuição de V e θ um vector de dimensão $r \times 1$ $(r \geq k)$ de parâmetros desconhecidos associados às variáveis aleatórias incluídas em X e também outros parâmetros específicos ao modelo probabilístico de T. As funções de distribuição e de densidade de t condicionais em x e em v são, respectivamente

$$F(t|x, v, \theta), \tag{1}$$

$$f(t|x, v, \theta). \tag{2}$$

Com base nas funções de distribuição e de densidade é possível construir a função *hazard* ou função intensidade de falha condicional em x e v:

$$\lambda(t|x, v, \theta) = f(t|T > t, x, v, \theta) = \frac{f(t|x, v, \theta)}{1 - F(t|x, v, \theta)} \geq 0. \tag{3}$$

Pela definição (3), percebe-se que a função *hazard* representa a densidade de T condicional no facto de $T > t$. Uma outra forma de interpretar a função *hazard* é a de "probabilidade" instantânea de saída do estado no intervalo $[t; t + \Delta t)$ dado que

[3]No caso geral, X pode ser variável no tempo. No entanto, a análise desta situação está fora do âmbito do presente texto uma vez que no contexto do modelo tradicional de *hazard* proporcional os regressores têm que ser constantes ao longo do tempo.

se esteve no estado até ao instante t, i.e,

$$\lambda(t|x,v,\theta) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T > t, x, v, \theta)}{\Delta t} = \frac{f(t|x,v,\theta)}{1 - F(t|x,v,\theta)}. \quad (4)$$

Uma outra função utilizada neste tipo de análise é a função de sobrevivência, i.e., a probabilidade de uma duração ser superior a t condicional no facto de se ter observado x e v. Assim sendo,

$$S(t|x,v,\theta) = P(T > t|x,v,\theta) = 1 - F(t|x,v,\theta). \quad (5)$$

Notando que as funções *hazard* e sobrevivência estão relacionadas da seguinte forma:

$$\lambda(t|x,v,\theta) = -\frac{\partial \ln\left(S(t|x,v,\theta)\right)}{\partial t}. \quad (6)$$

Integrando a expressão (6) em ordem a t (tendo presente que $S(0|x,v,\theta) = 1$), obtém-se o seguinte resultado:

$$S(t|x,v,\theta) = \exp\left\{ -\int_0^t \lambda(u|x,v,\theta)du \right\}. \quad (7)$$

Uma última função que é importante referir é a função *hazard* integrada. Apesar de não ter qualquer interpretação probabilística, esta função desempenha um papel muito importante no desenvolvimento de testes de especificação, uma vez que a generalidade dos erros deste tipo de modelos são construídos com base nesta mesma função

$$\Lambda(t|x,v,\theta) = \int_0^t \lambda(u|x,v,\theta)du. \quad (8)$$

Uma particularidade importante desta função é o facto de ela ter distribuição exponencial unitária[4], i.e.,

$$\Lambda(t|x, v, \theta) \sim Exp(1). \tag{9}$$

Sabendo que a função *hazard* mede a "probabilidade" instantânea de sair do estado condicional no facto de que se esteve no estado até esse momento, importa saber se esta probabilidade se altera com o tempo que se está no estado. Quando a função *hazard* varia com t diz-se que existe dependência da duração e esta dependência pode ser classificada através da derivada parcial da função *hazard* em ordem a t da seguinte forma:

$$\frac{\partial \lambda (t|x, v, \theta)}{\partial t} \begin{cases} < 0 & \text{dependência da duração negativa,} \\ = 0 & \text{não existe dependência da duração,} \\ > 0 & \text{dependência da duração positiva.} \end{cases} \tag{10}$$

Quando existe dependência da duração negativa, significa que quanto mais tempo se está no estado menos provável é a saída do mesmo. No caso de existir dependência da duração positiva, quanto mais tempo se está no estado mais provável é a saída. Se não existe dependência da duração, então, a probabilidade de saída do estado é constante ao longo do tempo que se permanece no mesmo.

2.2 Modelo de *Hazard* Proporcional

Na sequência dos conceitos da secção anterior, podemos definir o modelo de *Hazard* Proporcional - Cox (1972). Um modelo de duração diz-se ser de *Hazard* Proporcional se for possível factorizar a função *hazard* em duas funções, a primeira

[4]Veja-se Apêndice 2B

apenas função de t e a segunda apenas função de X e V ou seja,

$$\lambda(t|x,v,\theta)=\lambda_0(t)f(x,v,\beta). \tag{11}$$

Pela equação anterior percebe-se que, nos modelos de *hazard* proporcional, o efeito dos regressores sobre a função *hazard* acontece de forma multiplicativa através da função $f(x,v,\beta)$. Quando a função $f(x,v,\beta)$ é igual a 1, a função *hazard* reduz-se apenas ao elemento $\lambda_0(t)$, o qual é denominado por *baseline hazard*. Para que o modelo esteja totalmente especificado é ainda necessário definir as formas funcionais de $\lambda_0(t)$[5], $f(x,v,\beta)$ e a função $G(v)$[6]. Resumidamente, na especificação de um modelo de *Hazard* Proporcional é necessário fazer 4 hipóteses distintas:

1. Proporcionalidade da *Hazard*;
2. Forma funcional de $f(x,v,\beta)$;
3. Forma funcional de $\lambda_0(t)$;
4. Função $G(v)$.

De um modo geral, a incorrecta especificação de uma destas hipóteses conduzirá a estimadores inconsistentes para θ. Heckman e Singer (1984) mostraram ainda que, a incorrecta especificação de $G(v)$ não só torna os estimadores usuais inconsistentes para θ mas também os parâmetros que permitem medir a dependência da

[5]Veremos mais à frente que, sob determinadas hipóteses, a especificação desta função não é necessária para a estimação consistente dos parâmetros associados aos regressores X, i.e., os parâmetros incluídos no vector β.

[6]Sob as hipóteses de *hazard* proporcional e $E[V]=1$ é possível estimar os parâmetros de interesse sem especificar $G(v)$, Elbers e Ridder (1982).

duração vêm enviesados para baixo. A existência de heterogeneidade negligenciada mal modelada dá origem a que a dependência da duração seja mais negativa.

2.3 Censura

Em muitos casos, algumas observações são incompletas, i.e., não se observa o início, ou não se observa o fim ou nem se observa o início nem o fim do fenómeno que se está a estudar. Nestas condições, diz-se que a observação está censurada. A existência de censura prende-se essencialmente com o facto de a recolha da amostra não se poder prolongar indefinidamente no tempo, pelo que é necessário terminá-la em algum momento fazendo com que as durações que ainda estão a decorrer naquele momento sejam censuradas. Este tipo de censura é conhecido como censura à direita. Uma outra situação, também ligada ao processo amostral, tem que ver com o facto de no momento em que se dá início à recolha da amostra algumas durações já estarem a decorrer e para estes casos não ser possível saber o momento exacto em que começaram. Este tipo de censura é conhecido como censura à esquerda. Em dados económicos, o tipo de censura mais frequente é a censura à direita. De um ponto de vista formal, a variável que é efectivamente observada é a seguinte[7]:

$$T^* = \min\{T, C\}, \tag{12}$$

T representa a duração e C representa o ponto de censura. Sempre que $T < C$ observa-se $t^* = t$, se $T \geqslant C$, então, o que se observa é $t^* = c$. Assumindo que as

[7]Para uma discussão detalhada sobre os vários tipos de censura veja-se Cox (1984) ou Jenkins (2003).

variáveis T e C são independentes e também que o processo que gera a censura nos dados não é informativo sobre T, é possível estimar θ utilizando simultaneamente as observações censuradas e as não censuradas.

2.4 Amostragem

Apesar de existirem vários métodos de amostragem, há dois que se destacam: amostragem do fluxo e amostragem do stock. Diz-se que uma amostra é do fluxo quando a partir de um dado momento - t_0 - se registam todas as durações que tenham início após esse momento. Neste caso, a amostra pode-se considerar aleatória, uma vez que, desde que o fluxo de novas durações seja constante, a probabilidade de se recolherem durações curtas ou durações longas apenas está relacionada com distribuição deste tipo de durações na população. Desta forma, a contribuição individual para a verosimilhança é a habitual. Para as observações completas temos,

$$f(t|x,v,\theta), \tag{13}$$

e para as observações censuradas,

$$S(t|x,v,\theta). \tag{14}$$

Diz-se que uma amostra é do stock quando num dado momento - t_0 - é recolhida uma amostra de durações que estão a decorrer em t_0, ou seja, recolhe-se uma amostra do stock de durações. Ao contrário da amostragem do fluxo, este processo de amostragem não é aleatório. A razão para a amostragem do stock não ser aleatória tem que ver com o facto de as durações longas terem maior probabilidade de estarem na amostra do que as durações curtas quando comparado com a distribuição deste tipo de durações na população. Esta forma de amostragem é enviesada pelo tamanho (*"length biased sampling"*). Sob as hipóteses de estacionaridade da distribuição de durações ao longo do tempo e fluxo constante de novas durações ao longo do tempo, as contribuições individuais para a verosimilhança são, respectivamente, para as observações completas e para as observações censuradas,

$$\frac{tf(t|x,v,\theta)}{E(t|x,v,\theta)}, \tag{15}$$

$$\frac{S(t|x,v,\theta)}{E(t|x,v,\theta)}. \tag{16}$$

As duas expressões anteriores são válidas tanto na situação em que se conhece o tempo que já decorreu, como nos casos em que apenas se observa o tempo que falta decorrer até e após t_0, respectivamente. Uma outra situação de amostragem do stock é aquela em que no momento t_0 se recolhe uma amostra das durações que estão a decorrer e, para cada elemento da amostra, se sabe o tempo que já decorreu até t_0 e depois se observam as mesmas durações até t_1. Neste caso, a amostra continua a ser uma amostra não aleatória, uma vez que é uma amostra do stock. No entanto, este processo de amostragem tem a vantagem de não ser necessário impôr as hipóteses relativamente restritivas do caso anterior (população constante e estacionaridade do processo) para deduzir as expressões da verosimilhança. Se t^E representar o tempo decorrido até t_0, t^I o tempo entre t_1 e t_0 e t^O o tempo após t_0 para os casos em que a duração termina no intervalo (t_0, t_1), as contribuições para a verosimilhança são as seguintes: quando se conhece a totalidade da duração, tem-se,

$$\frac{S(t^E|x,v,\theta) - S(t^O|x,v,\theta)}{S(t^E|x,v,\theta)}, \tag{17}$$

quando não se conhece a totalidade da duração, tem-se,

$$\frac{S(t^E|x,v,\theta) - S(t^I|x,v,\theta)}{S(t^E|x,v,\theta)}. \tag{18}$$

Este tipo de amostragem é por vezes referido como amostragem por intervalo (*"observation over fixed interval"* ou *"stock sampling with follow-up"*).

2.5 Exemplos de modelos de *Hazard* Proporcional

Como exemplos de modelos de *hazard* proporcional escolhemos 3 casos particulares que nos permitirão tecer algumas considerações. Uma vez que todos os exemplos são casos de modelos de *hazard* proporcional, a primeira hipótese de especificação está determinada, isto é, a *hazard* é proporcional. Em relação à forma funcional de $f(x,v,\beta)$, apesar de existirem imensas alternativas, optámos pela mais comum, i.e., $f(x,v,\beta) = \exp(x\beta)v$. Assim sendo, o que diferencia os três exemplos que se seguem são as escolhas feitas para $\lambda_0(t)$ e para $G(v)$.

- Exemplo 1

O primeiro exemplo é a função Weibull com heterogeneidade negligenciada tipo Gama. Neste caso, $\lambda_0(t) = \alpha t^{\alpha-1}$ e $v \sim Gama(\delta,\delta)$,

$$\begin{aligned} \lambda(t|x,v,\theta) &= \alpha t^{\alpha-1}\exp(x\beta)\,v, \\ v &\sim Gama(\delta,\delta). \end{aligned} \tag{19}$$

Neste exemplo, quando se elimina a heterogeneidade negligenciada através da integração de v, a função que se obtém deixa de ser do tipo de *Hazard* Proporcional,

$$E_v\left[\lambda(t|x,v,\theta)\right] = \int_v \alpha t^{\alpha-1}\exp(x\beta)v dG(v) = \frac{\alpha t^{\alpha-1}\exp(x\beta)}{\left[1+\frac{t^\alpha}{\delta}\exp(x\beta)\right]}. \tag{20}$$

No entanto, condicional em x e em v a proporcionalidade existe. Este exemplo é particularmente interessante, porque mostra que a hipótese da Proporcionalidade da *Hazard* pode esconder-se através das hipóteses que se façam sobre a forma da heterogeneidade negligenciada. Fazendo $\alpha = 1 \implies \lambda_0(t) = 1$ e $G(v) = 1$ quando $v = 1$ e $G(v) = 0$ quando $v \neq 1$, obtém-se o modelo de duração mais simples que existe, isto é, o modelo exponencial,

$$\begin{gathered} \lambda(t|x,v,\theta) = \exp(x\beta)v, \\ G(v) = \begin{cases} 1 & se\ v = 1, \\ 0 & se\ v \neq 1. \end{cases} \end{gathered} \tag{21}$$

- Exemplo 2

O segundo exemplo é a *hazard* constante por troços[8]. Suponha-se que a variável T é dividida em k intervalos: $[0;t_1]$, $(t_1;t_2]$, $(t_2;t_3]$, ..., $(t_{k-1};+\infty)$, sendo razoável assumir que a função hazard é constante nesses intervalos e que não existe heterogeneidade negligenciada, o modelo em causa é:

$$\begin{gathered} \lambda(t|x,v,\theta) = \begin{cases} \lambda_1\exp(x\beta)v & se\ 0 \leq t \leq t_1, \\ \lambda_2\exp(x\beta)v & se\ t_1 < t \leq t_2, \\ (...) & \\ \lambda_k\exp(x\beta)v & se\ t > t_{k-1}. \end{cases} \\ G(v) = \begin{cases} 1 & se\ v = 1, \\ 0 & se\ v \neq 1. \end{cases} \end{gathered} \tag{22}$$

[8] Em inglês, "Piecewise constant".

Esta função tem a vantagem de ser bastante flexível quanto à forma da *hazard*. Muitos dos trabalhos aplicados que são feitos hoje em dia utilizam este tipo de especificação.

- Exemplo 3

O terceiro e último exemplo, é o modelo de Cox. Cox (1972), utilizando a propriedade da proporcionalidade da *hazard*, mostrou que era possível estimar os parâmetros β de um modelo de *hazard* proporcional sem ser necessário especificar a função "*baseline hazard*" - $\lambda_0(t)$.

Seja $\{(T_i^*, X_i)\ i = 1, ..., N\}$ uma amostra aleatória de dimensão N, na qual podem existir observações censuradas à direita. Ordene-se esta amostra por ordem crescente, i.e, $t_1^* < t_2^* < ... < t_{n-1}^* < t_n^*$. Sempre que uma duração termina ou é censurada existe um grupo de observações que também podiam ter terminado ou ter sido censuradas, seja $R(t)$ esse grupo. Sabendo que num dado momento terminou uma duração ou que uma duração foi censurada, podemos calcular a probabilidade de essa duração terminar condicional no facto de no grupo $R(t)$ ter terminado apenas uma observação ou calcular a probabilidade de essa duração ser censurada condicional no facto de apenas uma das durações pertencentes ao grupo $R(t)$ ter sido censurada. Estas probabilidades são, respectivamente,

$$\frac{\lambda(t_j^*|x_j, v, \theta)}{\sum_{t_i^* \in R(t)} \lambda(t_j^*|x_i, v, \theta)} \tag{23}$$

$$1 \tag{24}$$

Utilizando a propriedade da proporcionalidade, a expressão (23) simplifica-se da seguinte forma:

$$\frac{\lambda_0(t_j^*)f(x_j,v,\beta)}{\sum_{t_i^*\in R(t)}\lambda_0(t_j^*)f(x_i,v,\beta)}=\frac{f(x_j,v,\beta)}{\sum_{t_i^*\in R(t)}f(x_i,v,\beta)} \tag{25}$$

O elemento essencial neste modelo é o facto de a função $\lambda_0(t)$ ser eliminada da expressão (23). Ao eliminar $\lambda_0(t)$ de (23) é possível estimar β sem conhecer ou ter que especificar $\lambda_0(t)$. Se, por qualquer motivo, o objectivo da análise recair sobre a função $\lambda_0(t)$ então será necessário estimar esta função.

Um aspecto que é importante referir acerca deste modelo é que o método de estimação não é Máxima Verosimilhança mas sim Verosimilhança Parcial. Ainda assim, algumas propriedades dos estimadores de Máxima Verosimilhança também são válidas para este estimador. Em particular, a distribuição assimptótica é a mesma para os dois estimadores.

2.6 Estimação e inferência

Seja $\{(T_i^*,X_i)\ i=1,...,N\}$ uma amostra aleatória do fluxo e d_i uma variável binária que toma o valor 1 quando a observação i é censurada e 0 no caso contrário. Nestas condições, a contribuição individual para a verosimilhança é a seguinte:

$$l(\theta|t_i^*,x_i)=\int_v\left\{[f(t_i^*|x_i,v,\theta)]^{1-d_i}\,[S(t_i^*|x_i,v,\theta)]^{d_i}\right\}dG(v), \tag{26}$$

ou, de forma equivalente,

$$l(\theta|t_i^*, x_i) = \int_v \left\{ [\lambda(t_i^*|x_i, v, \theta)]^{1-d_i} [S(t_i^*|x_i, v, \theta)] \right\} dG(v). \quad (27)$$

O estimador de máxima verosimilhança de θ - $\widehat{\theta}_{MV}$ - é o que resulta do seguinte problema de optimização:

$$\widehat{\theta}_{MV} : \arg\max_{\theta} \ln \left\{ \prod_{i=1}^{N} \int_v \left\{ [\lambda(t_i^*|x_i, v, \theta)]^{1-d_i} [S(t_i^*|x_i, v, \theta)] \right\} dG(v) \right\}. \quad (28)$$

Sob determinadas condições de regularidade,

$$\sqrt{N}(\widehat{\theta}_{MV} - \theta) \overset{a}{\sim} N(0; [I(\widehat{\theta}_{MV})]^{-1}). \quad (29)$$

O caso de estimação por Verosimilhança Parcial é bastante similar à estimação por Máxima Verosimilhança uma vez que também é a solução de um determinado problema de optimização, em particular,

$$\widehat{\beta}_{VP} : \arg\max_{\beta} \prod_{j=1}^{N} \left[\frac{f(x_j, v, \beta)}{\sum_{t_i^* \in R(t)} f(x_i, v, \beta)} \right]^{1-d_j}. \quad (30)$$

Tsiatis (1981) demonstrou a consistência e a normalidade assimptótica do estimador anterior,

$$\sqrt{N}(\widehat{\beta}_{VP} - \beta) \overset{a}{\sim} N(0; [I(\widehat{\beta}_{VP})]^{-1}). \quad (31)$$

2.7 Erros e Resíduos

Na maioria dos modelos econométricos os resíduos têm um papel muito importante quando se pretende testar a especificação feita para o modelo em causa. De uma forma genérica, os erros do modelo são definidos como a diferença entre a variável dependente e o valor esperado condicional ou um quantil condicional. Os resíduos do modelo são a diferença entre a variável dependente e a estimativa do valor esperado condicional ou de um quantil condicional. No caso dos modelos de duração, os conceitos de erros mais utilizados não utilizam directamente a variável dependente, mas sim uma função monótona da mesma cujas propriedades são conhecidas e gerais. A transformação em causa é a função *hazard* integrada. Assim sendo, podemos definir o erro de um modelo de duração da seguinte forma,

$$e = \Lambda(t|x, v, \theta). \tag{32}$$

Na literatura estes tipo de erros é conhecido como erros generalizados ou erros no sentido de Cox e Snell. Quando não existe censura, estes erros têm a particularidade de terem distribuição exponencial unitária. No entanto, quando existe censura, as suas propriedades estatísticas não são conhecidas.

Dois resultados particularmente interessantes acerca das propriedades da *hazard* integrada foram apresentados por Chesher e Lancaster (1985). Primeiro resultado: apesar de e não ter distribuição exponencial quando existe censura, adicionando uma variável exponencial unitária a e nos casos em que estes dizem respeito a observações censuradas, a variável resultante - $\widetilde{e}$ - tem distribuição exponencial unitária,

$$\widetilde{e} = e + du \sim Exp(1). \tag{33}$$

A variável D tem o significado habitual, isto é, toma o valor 0 quando a observação não é censurada e o valor 1 caso contrário. A variável U é uma variável aleatória simulada com distribuição $Exp(1)$.

Segundo resultado: é possível criar uma nova variável a partir de e cujo valor esperado é 1 e variância $1 - P(D = 1)$,

$$e^* = e + d. \tag{34}$$

Estes resultados são particularmente interessantes porque permitem a realização de testes de especificação. Para uma demonstração das propriedades aqui referidas veja-se o apêndice 2B.

3 Testes de Especificação ao Modelo de *Hazard* Proporcional

Nesta secção apresentamos alguns testes de especificação ao modelo de *Hazard* Proporcional. Por forma a tornar a exposição mais clara separámos os testes consoante estes sejam formais ou não formais. Os testes formais foram ainda divididos de acordo com a hipótese do modelo de *Hazard* Proporcional que pretendem testar.

Os testes não formais são essencialmente testes gráficos dos quais não será possível inferir estatisticamente quer sobre a má especificação quer sobre a origem da má especificação. No entanto, à semelhança do que acontece com outros modelos econométricos, este tipo de testes poderá ajudar a desvendar alguns erros que outros testes poderão não detectar.

Em termos de testes formais optámos por os separar em cinco sub-secções diferentes: as quatro primeiras correspondem às quatro hipóteses de especificação do modelo de *Hazard* Proporcional apresentadas na sub-secção 2.2 e a quinta sub-secção apresenta testes de especificação geral. No Apêndice 1 apresentamos as cinco formas de teste que são mencionadas ao longo do texto.

3.1 Testes não formais

Tal como nos modelos lineares, a análise visual de resíduos pode ser útil na detecção de problemas ou erros de especificação. Este tipo de análise pode ser feita de várias formas. No entanto, apenas iremos apresentar o método proposto por Chesher, Lancaster e Irish (1985) uma vez que em nossa opinião este método é o mais simples e também o mais geral.

Este teste tira partido das propriedades da função *hazard* integrada, nomeadamente, ter distribuição exponencial unitária no caso em que não existe censura e nos casos em que existe censura ser possível fazer uma transformação a esta mesma função que faça com que continue a ter uma distribuição exponencial unitária (veja-se Apêndice 2B).

Sejam $\Lambda(t|x,v,\theta)$ a função *hazard* integrada populacional e $\widehat{\Lambda}(t|x,v,\widehat{\theta})$ um estimador consistente da mesma função. Sabendo que na população $\Lambda(t|x,v,\theta)$ tem distribuição $Exp(1)$, se o modelo estiver bem especificado, então, a distribuição de $\widehat{\Lambda}(t|x,v,\widehat{\theta})$ deverá ser bastante próxima da distribuição $Exp(1)$. Estimando a distribuição amostral de $\widehat{\Lambda}(t|x,v,\widehat{\theta})$ através de um qualquer estimador não paramétrico (por exemplo o estimador Kaplan-Meier[9]) e fazendo a representação no plano cartesiano de $-\ln[\widehat{S}(\widehat{\Lambda}(t|x,v,\widehat{\theta}))]$ com $\widehat{\Lambda}(t|x,v,\widehat{\theta})$, se o modelo tiver sido correctamente especificado, então, os vários pontos deverão estar todos na bissectriz do mesmo plano. Uma outra representação gráfica igualmente válida e eventualmente mais conveniente é representar $\widehat{S}(\widehat{\Lambda}(t|x,v,\widehat{\theta}))$ e $\exp\{-\widehat{\Lambda}(t|x,v,\widehat{\theta})\}$. A vantagem desta representação é a de os vários valores estarem limitados ao intervalo $[0,1]$.

A existência de censura faz com que a função $\Lambda(t|x,v,\theta)$ deixe de ter uma distribuição conhecida. Logo, não existe razão formal para que $-\ln[\widehat{S}(\widehat{\Lambda}(t|x,v,\widehat{\theta}))]$ esteja muito próximo de $\widehat{\Lambda}(t|x,v,\widehat{\theta})$, ou de forma equivalente, $\widehat{S}(\widehat{\Lambda}(t|x,v,\widehat{\theta}))$ de $\exp\{-\widehat{\Lambda}(t|x,v,\widehat{\theta})\}$. Uma forma de recuperar este teste passa por utilizar o resultado da equação (33), i.e., adicionar uma variável exponencial unitária simu-

[9] Para uma exposição detalhada deste estimador veja-se Cox e Oakes (1984) ou Lancaster (1990).

lada a $\Lambda(t|x,v,\theta)$. Fazendo isto, se o modelo estiver correctamente especificado, $-\ln[\widehat{S}(\widehat{\Lambda}(t|x,v,\widehat{\theta}))]$ e $\widehat{\Lambda}(t|x,v,\widehat{\theta})$ (ou $\widehat{S}(\widehat{\Lambda}(t|x,v,\widehat{\theta}))$ e $\exp\{-\widehat{\Lambda}(t|x,v,\widehat{\theta})\}$) deverão estar muito próximas.

3.2 Testes Formais

3.2.1 Testes à Proporcionalidade da função *Hazard*

Uma das características do modelo de *Hazard* Proporcional é o facto de nenhum dos regressores poder variar no tempo, i.e., todos os regressores devem ser constantes. Dada esta restrição, Cox (1972) propôs que se incluissem um ou mais regressores variáveis no tempo - $z(t)$ - e, um teste sobre a nulidade dos parâmetros associados a $z(t)$ seria um teste à Proporcionalidade da *Hazard*,

$$\lambda(t|x,v,\theta) = \lambda_0(t)g(x,v,\beta,z(t),\gamma), \tag{35}$$

$$H_0 : \gamma = 0.$$

Esta hipótese pode ser testada utilizando qualquer uma das cinco metodologias de teste. McCall (1994) propôs uma extensão deste teste de forma a permitir que se controlasse para a existência de heterogeneidade negligenciada e para o facto de, em aplicações empíricas, os dados disponíveis estarem agrupados. A vantagem deste novo teste é permitir isolar factores como heterogeneidade negligenciada ou discretização dos dados. Seja,

$$\lambda(t|x,v,\theta) = \lambda_0(t)\exp(x(t)'\beta(t))v. \tag{36}$$

Suponhamos que a variável T está agrupada em intervalos e que os valores de $X(t)$ são observados apenas no ínicio de cada intervalo. Desta forma, $K = k$ se $T \in (K-1; K]$ e $X_K = X(t-1)$, o que dá origem à seguinte função:

$$P(K = k | K > k-1, x_1, ..., x_k, v) = 1 - \exp(-\exp(\alpha_k + x_k' \beta_k) v), \qquad (37)$$

onde[10],

$$\alpha_k = \ln \left\{ \int_{k-1}^{k} \lambda_0(t) dt \right\}, \quad k = 1, ..., D, \qquad (38)$$

$$\beta_k = \int_{k-1}^{k} \beta(t) dt, \quad k = 1, ..., D. \qquad (39)$$

Testar a proporcionalidade da *hazard* é equivalente a testar

$$H_0 : \beta_1 = ... = \beta_D.$$

Mais uma vez, este teste pode ser feito utilizando qualquer uma das cinco formas de teste.

3.2.2 Testes à forma funcional de $f(x, v, \beta)$

Em aplicações empíricas, é usual considerar $f(x, v, \beta) = f(g(x, \beta), v) = \exp(x'\beta)$. No entanto, esta hipótese pode não ser válida. Uma característica desta especificação é ter por base três hipóteses diferentes, (i) não existe heterogeneidade negligenciada, ou seja, $Var(v) = 0$. Por agora, vamos manter esta hipótese como sendo válida; (ii) considera-se que $f(\cdot) = \exp(\cdot)$; (iii) os regressores X e os parâmetros β estão rela-

[10] Esta expressão pode ser vista com sendo a expressão de um modelo "piecewise constant" no qual se considera um troço para cada duração possível.

cionados linearmente - $g(x, \beta) = x'\beta$. Para testar estas duas últimas hipóteses, em separado ou em conjunto, vamos utilizar uma especificação que permita *encaixar* ambas hipóteses. Seguindo uma proposta de Ramsey (1969) e outra de Wooldridge (1992) podemos escrever $f(x, v, \beta)$ da seguinte forma[11]:

$$f(x, v, \beta) = [1 + \delta g(x, \beta)]^{\frac{1}{\delta}} \simeq [1 + \delta(x'\beta + \rho(x'\beta)^2)]^{\frac{1}{\delta}}, \quad \delta \in \Re. \tag{40}$$

O que é interessante nesta especificação é o facto de, no limite, quando $\delta = 0$ e $\rho = 0$, se obter $f(x, v, \beta) = \exp(x'\beta)$. Desta forma, testar $f(x, v, \beta) = \exp(x'\beta)$, é equivalente a testar $\delta = 0$ e $\rho = 0$, teste este que pode ser feito utilizando qualquer uma das 5 formas de teste. O problema de utilizar o teste de Wald ou o teste LR é o facto de ser necessário estimar δ, o que, apesar de ser teoricamente possível, para alguns valores de δ iria criar restrições nos outros parâmetros (ρ e β). Assim sendo, em nossa opinião, os testes que mais se adequam neste caso são os testes LM/CM e os testes $C(\alpha)$ uma vez que apenas é necessário estimar o modelo sob a hipótese nula, i.e., considerando $f(x, v, \beta) = \exp(x'\beta)$. No caso em que se testam as duas hipóteses em simultâneo os testes LM e $C(\alpha)$ são equivalentes. Um aspecto que é preciso realçar é o facto de não ser necessário testar as duas hipóteses em simultâneo. Neste caso, e de acordo com Neyman (1959) e Jaggia e Trivedi (1994), o teste $C(\alpha)$ tem vantagens sobre o teste LM uma vez que a realização do teste $C(\alpha)$ sobre uma das hipóteses não implica a validade da outra.

[11]Uma vez que $g(x, \beta)$ é desconhecida, e seguindo a sugestão de Ramsey, aproximámos esta função por um polinómio de 2º grau.

3.2.3 Testes à forma funcional de $\lambda_0(t)$

A forma funcional de $\lambda_0(t)$ é um elemento fundamental neste tipo de modelação. No entanto, apenas é necessário especificar esta função quando se opta por uma abordagem de modelação completamente paramétrica, i.e., quando se utiliza o modelo de Cox (abordagem semi-paramétrica) não é necessário definir $\lambda_0(t)$. Se se opta por uma abordagem paramétrica, a escolha de $\lambda_0(t)$ é bastante importante pelo que é necessário testar essa mesma escolha.

Sejam $\lambda_{0,1}(t)$ e $\lambda_{0,2}(t)$ duas especificações alternativas para $\lambda_0(t)$. Supondo que desejamos testar $\lambda_{0,1}(t)$ contra $\lambda_{0,2}(t)$, consideremos três situações distintas: 1) a função $\lambda_{0,1}(t)$ é um caso particular de $\lambda_{0,2}(t)$ e como tal, $\lambda_{0,1}(t)$ obtém-se impondo restrições sobre algum(ns) parâmetro(s) de $\lambda_{0,2}(t)$. Neste caso, diz-se que $\lambda_{0,1}(t)$ está *encaixada* em $\lambda_{0,2}(t)$; 2) a função $\lambda_{0,1}(t)$ não é um caso particular de $\lambda_{0,2}(t)$, mas, existe uma terceira função, $\lambda_{0,3}(t)$, que tem como casos particulares $\lambda_{0,1}(t)$ e $\lambda_{0,2}(t)$. Neste caso diz-se que $\lambda_{0,1}(t)$ e $\lambda_{0,2}(t)$ estão *encaixadas* em $\lambda_{0,3}(t)$; 3) a função $\lambda_{0,1}(t)$ não é um caso particular de $\lambda_{0,2}(t)$ nem tão pouco existe uma terceira função, $\lambda_{0,3}(t)$, que tenha $\lambda_{0,1}(t)$ e $\lambda_{0,2}(t)$ como casos particulares. Neste caso, diz-se que $\lambda_{0,1}(t)$ e $\lambda_{0,2}(t)$ não estão *encaixadas.*

Como nas duas primeiras situações o modelo ou os modelos a testar são casos particulares de modelos mais gerais, a realização deste teste é relativamente simples, ou seja, nestes casos basta testar a validade das restrições que são necessárias impôr nos modelos mais gerais para que se obtenha o modelo mais restrito. De um modo geral, estas restrições podem ser testadas utilizando qualquer uma das cinco forma de teste apresentadas no Apêndice 1A. Por exemplo, consideremos os modelos Exponencial e Weibull. No modelo Exponencial tem-se $\lambda_0(t) = 1$ e no modelo Weibull tem-se $\lambda_0(t) = \gamma t^{\gamma-1}$, nestas condições, testar o modelo Exponencial contra o modelo Weibull é equivalente a testar $\gamma = 1$. Um problema desta abordagem é o de se assumir, implicitamente, que o modelo geral está correcto quando se rejeita o modelo restrito.

A situação 3), situação em que nenhum dos modelos a testar é caso particular do outro, nem tão pouco existe um terceiro modelo que tenha ambos modelos como casos particulares, é algo mais complexa. Nestas situações é necessário recorrer a um tipo de testes conhecido na literatura como teste de hipóteses não *encaixadas*. O teste que vamos apresentar é um teste proposto por Santos Silva (2001) que tem por base uma ideia de Cox (1961). Esta ideia, consiste em *encaixar* artificialmente os dois modelos a testar. Utilizando esta abordagem, os modelos que resultam de escolher $\lambda_{0,1}(t)$ ou $\lambda_{0,2}(t)$ podem ser *encaixados* artificialmente da seguinte forma:

$$\frac{\left[(1-\alpha)\left[\lambda_1(t)\right]^{1-d}\exp\left\{-\Lambda_1(t)\right\}+\alpha\left[\lambda_2(t)\right]^{1-d}\exp\left\{-\Lambda_2(t)\right\}\right]^{\frac{1}{\rho}}}{\int_0^{\infty}\left[(1-\alpha)\left[\lambda_1(t)\right]^{1-d}\exp\left\{-\Lambda_1(t)\right\}+\alpha\left[\lambda_2(t)\right]^{1-d}\exp\left\{-\Lambda_2(t)\right\}\right]^{\frac{1}{\rho}}dt}. \tag{41}$$

Nesta expressão, $\lambda_1(t) = \lambda_{0,1}(t) f(x, v, \beta)$ e $\lambda_2(t) = \lambda_{0,2}(t) f(x, v, \beta)$. Para testar $\lambda_{0,1}(t)$ contra $\lambda_{0,2}(t)$ basta realizar o seguinte teste,

$$H_0 : \alpha = 0 \ (\textit{ou no caso contrário, } \alpha = 1). \tag{42}$$

Para efectuar este teste é ainda necessário fixar o valor de ρ. O único valor de ρ que faz com que o denominador de (41) seja eliminado é $\rho = 1$. O teste proposto por Cox (1961) tinha por base um valor de $\rho = 0$ o que fazia com que para efectuar o teste fosse necessário calcular o denominador de (41). Este teste tem a particularidade de poder manter as duas hipóteses, rejeitar ambas ou manter apenas uma delas. Para que este teste seja válido, é necessário que as funções $[\lambda_1(t)]^{1-d}\exp\{-\Lambda_1(t)\}$ e $[\lambda_2(t)]^{1-d}\exp\{-\Lambda_2(t)\}$ sejam superiores a 0 e inferiores a ∞ em todo o seu domínio.

Este teste apenas pode ser feito através das metodologias de teste LM/CM ou $C(\alpha)$ uma vez que a restrição a testar está na fronteira do espaço parâmetro[12]. No apêndice 2A apresentamos a condição de momentos referente a este teste.

3.2.4 Testes à existência de heterogeneidade negligenciada

Uma situação recorrente em modelos microeconométricos é a existência de heterogeneidade negligenciada, i.e., caraterísticas individuais que por não serem observadas não são levadas em conta no modelo. As consequências desta situação em modelos de duração são de dois tipos, (i) por um lado o estimador deixa de ser consistente para os parâmetros de interesse e, (ii) por outro, cria um enviesamento na forma da dependência da duração para dependências da duração mais negativas. Dadas as consequências já mencionadas torna-se necessário testar se existe heterogeneidade negligenciada. Chesher (1984) propôs uma forma geral de teste à presença de heterogeneidade negligenciada que passa por considerar uma aproximação da função de verosimilhança através de um polinómio de Taylor. Lancaster (1985), derivou este mesmo teste para modelos de duração na ausência de censura, e Jaggia (1997) generalizou o teste para presença de censura.

Seja

$$L_i(\theta|t,x,v) = \lambda(t|x,\theta,v)^{1-d}S(t|x,\theta,v) \tag{43}$$

[12] Veja-se Godfrey (1988).

a contribuição individual para a verosimilhança. Considere-se que $E[V] = 1$ e $Var[V] = \sigma_v^2$. Fazendo uma expansão em polinómio de Taylor desta mesma função na variável V em torno do seu valor esperado obtém-se,

$$\begin{aligned} L_i(\theta|t,x,v) \simeq\ & \lambda(t|x,\theta,1)^{1-d}S(t|x,\theta,1) - \\ & \lambda(t|x,\theta,1)^{1-d}S(t|x,\theta,1)(\Lambda(t|x,\theta,1)-(1-d))(v-1)+ \\ & \frac{1}{2}\lambda(t|x,\theta,1)^{1-d}S(t|x,\theta,1)(\Lambda(t|x,\theta,1)^2-\Lambda(t|x,\theta,1)(1-d))(v-1)^2. \end{aligned} \tag{44}$$

Calculando o valor esperado da função anterior na variável aleatória V, obtém-se a seguinte expressão:

$$\begin{aligned} E_V\{L(\theta|t,x,v)\} \simeq\ & \lambda(t|x,\theta,1)^{1-d}S(t|x,\theta,1) - \\ & \frac{\sigma_v^2}{2}\lambda(t|x,\theta,1)^{1-d}S(t|x,\theta,1)(\Lambda(t|x,\theta,1)^2-\Lambda(t|x,\theta,1)(1-d)). \end{aligned} \tag{45}$$

Nesta última expressão, σ_V^2 é um parâmetro adicional, pelo que testar a existência de heterogeneidade negligenciada é equivalente a testar

$$H_0 : \sigma_V^2 = 0. \tag{46}$$

À semelhança do teste anterior, e pelas mesmas razões, este teste apenas pode ser feito através das metodologias LM/CM ou $C(\alpha)$.

3.2.5 Testes de especificação geral

Até agora, todos os testes formais apresentados são testes em que a hipótese nula é sobre um determinado aspecto de especificação. No entanto, existem outros testes em que a hipótese nula não recai sobre nenhuma hipótese de especificação em particular mas sim sobre a especificação do modelo como um todo. Estes testes podem ser bastante úteis uma vez que permitem testar a correcta especificação do modelo através de um único teste e, no caso de não rejeição da hipótese nula (especificação correcta) tem-se uma boa indicação da validade do modelo. O principal problema destes testes vem do facto de, quando a hipótese nula é rejeitada, não se terem quaisquer indicações sobre a origem da má especificação. Neste tipo de testes é comum testar propriedades de algum tipo de erros, por exemplo, testar as propriedades dos erros apresentados na secção 2.7.

Seja $e(t)$ uma variável de erro e $m[e(t)]$ uma função de $e(t)$ tal que, quando o modelo está correctamente especificado, se verifica $E\{m[e(t)]\} = 0$. Desta forma, o teste a realizar é:

$$H_0 : E\{m[e(t)]\} = 0. \tag{47}$$

Este teste baseia-se essencialmente em testes de Momentos Condicionais e $C(\alpha)$. As várias definições de erros utilizadas neste trabalho estão descritas no capítulo 2 e as propriedades testáveis de cada uma das definições assim como outras condições de momentos podem ser encontradas nos Apêndices 2A e 2B.

4 Simulação de Monte Carlo

Todas as metodologias de teste mencionadas na secção anterior e apresentadas no apêndice 1A têm em comum o facto de as suas distribuições serem derivadas assimptoticamente. Logo, para que estes resultados sejam válidos é necessário que a amostra com que se trabalha seja suficientemente grande.

Nesta secção, iremos analisar através de simulação de Monte Carlo as propriedades de pequenas amostras, nomeadamente dimensão e potência, dos testes de especificação geral apresentados na secção anterior, tanto no contexto de estimação por Máxima Verosimilhança (Tauchen-Newey) como no contexto do modelo de Cox (Horowitz e Neumann).

4.1 Dimensão

A dimensão (α) de um teste estatístico define-se como a probabilidade de rejeitar uma hipótese nula quando esta é verdadeira[13]. A realização de qualquer teste pressupõe que a dimensão do mesmo seja fixada *a priori*. Neste caso, fala-se em dimensão nominal ou assimptótica (α_N). Uma vez que a distribuição em pequenas amostras não é necessariamente igual à distribuição assimptótica, o investigador realiza o teste com base numa dimensão real (α_R) que desconhece e que poderá ser significativamente diferente da dimensão nominal por ele fixada.

Nesta sub-secção iremos analisar alguns factores que podem influenciar a dimensão dos testes de especificação baseados em momentos condicionais no contexto dos

[13] Veja-se Casella, G. e Berger, R. (2002), "Statistical Inference", 2^{nd} edition, Duxburry Press para uma possível definição formal de dimensão de um teste estatístico.

dois métodos. Por forma a analisar o impacto de diferentes factores - dimensão da amostra, proporção de observações censuradas, perfil da função *hazard*, distribuição dos regressores, número de regressores, inclusão de variáveis irrelevantes e sobreparametrização de $\lambda_0(t)$ - realizámos o seguinte exercício de simulação:

- Três dimensões da amostra: $N_1 = 100$, $N_2 = 500$ *e* $N_3 = 1000$;
- Três níveis de censura: $\pi_1 = 0\%$, $\pi_2 = 20\%$ *e* $\pi_3 = 50\%$;
- Quatro conjuntos de regressores: $X_1 = \{N(1,2)\}$, $X_2 = \left\{\chi^2_{(1)}\right\}$, $X_3 = \{Bernoulli\,(p = 0.3)\}$ *e* $X_4 = \{X_1, X_2, X_3\}$;
- Três funções Weibull[14] com parâmetro γ diferente: $\gamma_1 = 0.5$, $\gamma_2 = 1.0$ *e* $\gamma_3 = 1.5$.

Por uma questão de simplicidade fizemos $f(x, \beta, v) = \exp(x'\beta) = \exp(1 + 1'X_r)$, $r = 1, 2, 3$ e 4. Para gerar o níveis de censura π_2 e π_3 considerámos que as durações tinham início no intervalo $[0, M)$, ou seja, o início de cada observação é determinado por uma variável aleatória uniforme - U - definida no intervalo $[0, M]$ e todas as durações superiores ou iguais a $M - U$ são censuradas à direita

$$T^* = \min\{T, M - U\}. \tag{48}$$

Desta forma, os níveis de censura π_2 e π_3 são determinados por M

$$M_j : \int_0^M \int_X \int_V S(M_j - u|x, v, \theta) dG(v)\, dH(x)\, u\frac{1}{M} = \pi_j, \quad j = 2, 3. \tag{49}$$

[14]Função Weibull - $f(t|x, v, \theta) = \gamma\lambda(x, v, \beta)^{\gamma} t^{\gamma-1} \exp\{-[\lambda(x, v, \beta)t]^{\gamma}\}$

O valor de M varia não só com π, mas também com o vector de regressores X e com o parâmetro γ da função Weibull. No Apêndice 3 apresentamos os vários valores de M para as várias combinações (π, X, γ)[15].

A realização de testes de momentos implica a escolha de funções de momentos e a performance dos testes não é inócua à escolha destas funções. No entanto, neste trabalho não iremos analisar esta questão e utilizaremos a mesma função de momentos utilizada por Horowitz e Neumann (1992) uma vez que estes autores referem que foi para esta função que obtiveram os melhores resultados. A condição de momentos escolhida é a seguinte[16]:

$$(2 - D) S [\Lambda (T^* | x, \theta, v)] - 1. \tag{50}$$

Uma vez que a performance dos testes de momentos condicionais tipo Tauchen-Newey depende do estimador utilizado para a matriz de variâncias-covariâncias dos estimadores iremos apresentar os resultados para os dois principais estimadores, i.e., producto externo do gradiente (OPG) e matriz Hesseana (H). Quando se utiliza a matriz Hesseana, a variância das condições de momentos pode não ser semi-definida positiva, o que significa que a estatística de teste pode ser negativa. Nestes casos, considerámos que a hipótese nula deveria ser rejeitada uma vez que sob a mesma hipótese, a estatística de teste teria que ser positiva. No caso em que se opta por utilizar a matriz OPG, o problema agora referido não se coloca, uma vez que por construção a matriz de variâncias das condições de momentos é sempre semi-definida

[15] Todos estes valores foram obtidos através de simulação de Monte Carlo.
[16] Veja-se o Anexo 2B para mais detalhes sobre esta função.

positiva (resultado clássico de álgebra linear). Relativamente aos testes de momentos condicionais do modelo de Cox, os autores (Horowitz e Neumann) apresentam duas estatísticas de teste diferentes. Uma estatística de teste corrigida para pequenas amostras (CR) e uma outra não corrigida (NCR), de forma a perceber a importância da correcção optámos por apresentar os resultados para ambos casos. Importa ainda dizer que cada um dos resultados é obtido com base em 10'000 réplicas, o algoritmo de optimização é o algoritmo de Gauss-Newton, o qual utiliza derivadas analíticas da função logaritmo da verosimilhança, e o software utilizado foi o GAUSS[17]. Desta forma, passamos a apresentar os resultados obtidos.

Quadro 4.1.1 - Dimensão real dos testes para uma dimensão nominal de 5% quando o vector de regressores X_1 é utilizado.

		$N_1 = 100$			$N_2 = 500$			$N_3 = 1000$		
		π_1	π_2	π_3	π_1	π_2	π_3	π_1	π_2	π_3
γ_1	COX_{NCR}	0.150	0.153	0.135	0.076	0.067	0.065	0.062	0.057	0.053
	COX_{CR}	0.051	0.067	0.065	0.052	0.053	0.051	0.049	0.048	0.048
	CM_H	0.479	0.457	0.522	0.199	0.150	0.152	0.128	0.097	0.092
	CM_{OPG}	0.078	0.085	0.087	0.060	0.056	0.058	0.053	0.054	0.052
γ_2	COX_{NCR}	0.157	0.139	0.127	0.077	0.066	0.065	0.064	0.057	0.059
	COX_{CR}	0.041	0.054	0.054	0.051	0.053	0.053	0.049	0.052	0.051
	CM_H^+	0.074	0.096	0.169	0.054	0.059	0.085	0.053	0.056	0.069
	CM_H	0.478	0.486	0.552	0.201	0.176	0.162	0.130	0.114	0.104
	CM_{OPG}^+	0.079	0.081	0.078	0.054	0.053	0.058	0.057	0.055	0.057
	CM_{OPG}	0.080	0.086	0.095	0.060	0.059	0.063	0.054	0.054	0.056
γ_3	COX_{NCR}	0.214	0.142	0.112	0.098	0.070	0.062	0.076	0.061	0.057
	COX_{CR}	0.030	0.028	0.031	0.049	0.046	0.046	0.048	0.047	0.046
	CM_H	0.477	0.493	0.569	0.199	0.182	0.178	0.128	0.117	0.111
	CM_{OPG}	0.078	0.091	0.104	0.060	0.057	0.066	0.053	0.053	0.057

(+) Parâmetro γ não estimado.

[17]No cálculo do teste ao modelo de Cox utilizámos uma rotina em Gauss feita pelos autores do artigo.

Quadro 4.1.2 - Dimensão real dos testes para uma dimensão nominal de 5% quando o vector de regressores X_2 é utilizado.

		$N_1 = 100$			$N_2 = 500$			$N_3 = 1000$		
		π_1	π_2	π_3	π_1	π_2	π_3	π_1	π_2	π_3
γ_1	COX_{NCR}	0.202	0.180	0.137	0.081	0.067	0.062	0.065	0.057	0.048
	COX_{CR}	0.034	0.032	0.030	0.048	0.041	0.038	0.049	0.043	0.035
	CM_H	0.476	0.468	0.605	0.197	0.144	0.198	0.128	0.097	0.111
	CM_{OPG}	0.080	0.081	0.083	0.059	0.059	0.060	0.053	0.051	0.056
γ_2	COX_{NCR}	0.156	0.071	0.053	0.083	0.018	0.012	0.072	0.013	0.008
	COX_{CR}	0.022	0.009	0.008	0.043	0.007	0.005	0.045	0.007	0.005
	CM_H^+	0.073	0.103	0.180	0.055	0.060	0.093	0.054	0.057	0.072
	CM_H	0.476	0.474	0.595	0.199	0.160	0.178	0.129	0.106	0.104
	CM_{OPG}^+	0.082	0.077	0.082	0.054	0.053	0.059	0.058	0.054	0.060
	CM_{OPG}	0.081	0.083	0.088	0.060	0.057	0.053	0.053	0.053	0.055
γ_3	COX_{NCR}	0.126	0.029	0.020	0.081	0.004	0.002	0.073	0.002	0.001
	COX_{CR}	0.009	0.001	0.001	0.027	0.001	0.001	0.033	0.001	0.001
	CM_H	0.475	0.479	0.596	0.197	0.173	0.178	0.128	0.113	0.108
	CM_{OPG}	0.080	0.084	0.089	0.059	0.062	0.060	0.053	0.052	0.057

(+) Parâmetro γ não estimado.

Quadro 4.1.3 - Dimensão real dos testes para uma dimensão nominal de 5% quando o vector de regressores X_3 é utilizado.

		$N_1 = 100$			$N_2 = 500$			$N_3 = 1000$		
		π_1	π_2	π_3	π_1	π_2	π_3	π_1	π_2	π_3
γ_1	COX_{NCR}	0.538	0.541	0.460	0.162	0.163	0.127	0.106	0.105	0.082
	COX_{CR}	0.144	0.174	0.184	0.066	0.059	0.052	0.057	0.055	0.051
	CM_H	0.481	0.434	0.516	0.198	0.125	0.177	0.130	0.083	0.113
	CM_{OPG}	0.080	0.081	0.086	0.058	0.058	0.054	0.053	0.053	0.053
γ_2	COX_{NCR}	0.197	0.251	0.233	0.085	0.094	0.079	0.071	0.071	0.065
	COX_{CR}	0.096	0.057	0.056	0.064	0.052	0.050	0.061	0.053	0.051
	CM_H^+	0.071	0.106	0.207	0.055	0.064	0.104	0.054	0.057	0.079
	CM_H	0.476	0.460	0.523	0.201	0.156	0.149	0.135	0.105	0.093
	CM_{OPG}^+	0.081	0.076	0.067	0.054	0.055	0.052	0.058	0.051	0.049
	CM_{OPG}	0.080	0.083	0.084	0.058	0.056	0.059	0.053	0.054	0.052
γ_3	COX_{NCR}	0.180	0.198	0.200	0.076	0.080	0.075	0.061	0.064	0.059
	COX_{CR}	0.062	0.060	0.067	0.053	0.050	0.052	0.051	0.049	0.050
	CM_H	0.482	0.483	0.548	0.198	0.173	0.157	0.130	0.112	0.094
	CM_{OPG}	0.080	0.085	0.089	0.058	0.057	0.059	0.053	0.055	0.052

(+) Parâmetro γ não estimado.

Quadro 4.1.4 - Dimensão real dos testes para uma dimensão nominal de 5% quando o vector de regressores X_4 é utilizado.

		$N_1 = 100$			$N_2 = 500$			$N_3 = 1000$		
		π_1	π_2	π_3	π_1	π_2	π_3	π_1	π_2	π_3
γ_1	COX_{NCR}	0.167	0.139	0.110	0.076	0.060	0.055	0.065	0.051	0.049
	COX_{CR}	0.099	0.099	0.095	0.063	0.054	0.052	0.053	0.046	0.040
	CM_H	0.695	0.751	0.863	0.263	0.226	0.279	0.155	0.127	0.125
	CM_{OPG}	0.082	0.085	0.093	0.059	0.057	0.062	0.054	0.054	0.057
γ_2	COX_{NCR}	0.200	0.071	0.055	0.100	0.022	0.015	0.082	0.013	0.008
	COX_{CR}	0.046	0.018	0.018	0.050	0.009	0.009	0.051	0.008	0.006
	CM_H^+	0.063	0.084	0.147	0.052	0.055	0.067	0.052	0.052	0.061
	CM_H	0.694	0.752	0.874	0.265	0.252	0.282	0.161	0.146	0.146
	CM_{OPG}^+	0.114	0.112	0.126	0.060	0.059	0.066	0.061	0.052	0.060
	CM_{OPG}	0.082	0.087	0.103	0.060	0.059	0.064	0.054	0.053	0.058
γ_3	COX_{NCR}	0.204	0.034	0.023	0.123	0.004	0.002	0.102	0.002	0.001
	COX_{CR}	0.014	0.004	0.003	0.034	0.002	0.001	0.038	0.001	0.001
	CM_H	0.695	0.750	0.876	0.263	0.261	0.300	0.155	0.149	0.152
	CM_{OPG}	0.086	0.091	0.110	0.059	0.060	0.062	0.054	0.053	0.057

(+) Parâmetro γ não estimado.

Quadro 4.1.5 - Dimensão real dos testes para uma dimensão nominal de 5% quando a variável dependente é gerada utilizando o vector de regressores X_2 e os parâmetros são estimados utilizando o vector X_4.

		$N_1 = 100$			$N_2 = 500$			$N_3 = 1000$		
		π_1	π_2	π_3	π_1	π_2	π_3	π_1	π_2	π_3
γ_1	COX_{NCR}	0.188	0.167	0.118	0.083	0.067	0.062	0.067	0.057	0.050
	COX_{CR}	0.144	0.131	0.115	0.067	0.056	0.050	0.061	0.050	0.042
	CM_H	0.695	0.751	0.855	0.263	0.205	0.298	0.155	0.115	0.145
	CM_{OPG}	0.082	0.084	0.089	0.059	0.057	0.060	0.054	0.051	0.055
γ_2	COX_{NCR}	0.164	0.075	0.055	0.086	0.019	0.012	0.073	0.013	0.009
	COX_{CR}	0.063	0.019	0.017	0.054	0.009	0.007	0.053	0.007	0.006
	CM_H	0.694	0.753	0.865	0.264	0.225	0.283	0.157	0.129	0.139
	CM_{OPG}	0.082	0.087	0.096	0.059	0.060	0.060	0.054	0.054	0.057
γ_3	COX_{NCR}	0.128	0.025	0.020	0.084	0.004	0.002	0.076	0.002	0.001
	COX_{CR}	0.030	0.003	0.004	0.035	0.001	0.001	0.040	0.001	0.001
	CM_H	0.695	0.748	0.871	0.263	0.242	0.286	0.155	0.137	0.142
	CM_{OPG}	0.087	0.087	0.104	0.059	0.062	0.060	0.054	0.053	0.059

(+) Parâmetro γ não estimado.

Com base nos resultados apresentados tirámos as seguintes conclusões:

- Tanto a versão H do teste CM como a versão não corrigida do teste do modelo de Cox podem ser desastrosas. A principal razão para a versão H do teste CM apresentar resultados tão maus vem do facto de uma grande proporção das estatísticas de teste serem negativas (em alguns casos mais de 50%). No caso da versão não corrigida do teste do modelo de Cox, os autores já haviam alertado para a má performance do teste e por isso mesmo apresentaram uma versão corrigida para pequenas amostras, a qual apresentou resultados razoáveis;

- Mais uma vez se verifica que a dimensão da amostra é fundamental. Na grande maioria dos casos, à medida que a dimensão da amostra aumenta, a dimensão real do teste diminui. Existem situações no teste do modelo de Cox em que a dimensão real do teste se aproxima de zero quando a amostra aumenta.

- Contrariamente ao que seria de esperar, a existência de censura nos dados tem pouco impacto na performance dos testes, verificando-se apenas um ligeiro aumento da dimensão real. No caso do modelo de Cox, a existência de censura tende a fazer diminuir a dimensão real do teste;

- Como seria de esperar, os resultados são ligeiramente afectados pela presença de regressores assimétricos, principalmente no caso do modelo de Cox. Neste caso, a principal consequência da existência de regressores assimétricos não é a sobre-rejeição da hipótese nula, mas sim o contrário, ou seja, a sub-rejeição. Também importa referir que o impacto agora descrito apenas se verificou com a presença de regressores assimétricos contínuos (caso de X_2). Quando o regressor é assimétrico mas discreto (caso de X_3), os resultados aparentemente não são afectados;

- Um resultado interessante é a relação que se verifica entre a dimensão real e o parâmetro γ da função Weibull. Nos casos analisados, verificámos que, principalmente no modelo de Cox, quando o valor do parâmetro γ aumenta (a dependência da duração torna-se mais positiva) a dimensão real diminui. Este resultado pode estar relacionado com o facto de, no modelo Weibull, a variância de t diminuir com o parâmetro γ;

- Nas experiências realizadas, a sobre-parametrização de $\lambda_0(t)$ não teve um grande impacto no comportamento dos testes, apenas se notou um ligeiro aumento da frequência de rejeição;

- Neste exercício não ficou claro que a inclusão de variáveis irrelevantes tenha um efeito negativo na dimensão real dos testes, aliás, nas nossas experiências os resultados melhoraram quando foram acrescentadas variáveis irrelevantes;

- Para terminar, concluímos que este teste tem boas propriedades de dimensão mas, depende bastante da construção da estatística de teste. No caso dos testes CM a versão OPG é claramente a que apresenta melhores resultados, o que significa que deve ser utilizada em detrimento da versão H do mesmo teste (este resultado é algo contraditório com os resultados existentes na literatura, no entanto, face aos resultados que obtivemos, é esta a conclusão que se deve tirar). No que diz respeito ao modelo de Cox, ficou claro que se deve utilizar sempre a versão do teste que tem uma correcção para pequenas amostras.

4.2 Potência

A potência de um teste define-se como a probabilidade de rejeitar a hipótese nula quando esta é falsa[18]. Idealmente, a potência de um teste deverá ser 1, i.e., um teste deve rejeitar a hipótese nula quando esta é falsa em 100% dos casos.

Ao contrário do que acontece quando se pretende analisar a dimensão do teste, a análise da potência pode ser feita em várias direcções, i.e., podem existir inúmeras formas de má especificação. Embora, a origem da má especificação seja relevante para a potência do teste, dado o tema central deste trabalho, iremos apenas analisar a potência do teste contra a não verificação da hipótese de proporcionalidade da função *hazard*. Para realizarmos esta análise considerámos o seguinte exercício de simulação:

- Três dimensões da amostra: $N_1 = 100$, $N_2 = 500$ *e* $N_3 = 1000$;
- Três níveis de censura: $\pi_1 = 0\%$, $\pi_2 = 20\%$ *e* $\pi_3 = 50\%$;
- Três conjuntos de regressores: $X_4 = \left\{N(1,2),\ \chi^2_{(1)},\ Bernoulli\,(p = 0.3)\right\}$, $Z_1 = \left\{0.5\chi^2_{(1)}\right\}$ *e* $Z_2 = \{Bernoulli\,(p = 0.5)\}$;
- Três funções Weibull com parâmetro γ função de Z_j: $\gamma(Z) = 1 + \tau(-1 + Z)$ e $\tau_1 = 0.1$, $\tau_2 = 0.5$ *e* $\tau_3 = 1$ (para $\tau = 0$ o modelo reduz-se ao modelo exponencial);

A simulação de observações censuradas foi feita do mesmo modo que anteriormente (os valores de M para as combinações (π, τ, Z) são apresentados no apêndice

[18] Veja-se Casella, G. e Berger, R. (2002), "Statistical Inference", 2^{nd} edition, Duxburry Press para uma possível definição formal de potência de um teste estatístico.

3A). No caso de estimação por Máxima Verosimilhança utilizámos um modelo Weibull. Mais uma vez, os resultados foram obtidos com base em 10'000 réplicas, o algoritmo de optimização utilizado foi o algoritmo de Gauss-Newton e o software utilizado foi o Gauss. Uma vez descrito o exercício de simulação, passamos a apresentar os resultados obtidos.

Quadro 4.2.1 - Potência dos testes quando o parâmetro γ é função do vector de regressores Z_1.

		$N_1 = 100$			$N_2 = 500$			$N_3 = 1000$		
		π_1	π_2	π_3	π_1	π_2	π_3	π_1	π_2	π_3
τ_1	COX_{NCR}	0.202	0.075	0.060	0.105	0.020	0.015	0.083	0.013	0.008
	COX_{CR}	0.042	0.020	0.022	0.045	0.009	0.007	0.046	0.008	0.005
	CM_H	0.689	0.759	0.871	0.251	0.243	0.281	0.145	0.136	0.139
	CM_{OPG}	0.080	0.083	0.108	0.057	0.058	0.062	0.050	0.053	0.052
τ_2	COX_{NCR}	0.254	0.121	0.106	0.200	0.067	0.060	0.217	0.065	0.058
	COX_{CR}	0.043	0.019	0.018	0.040	0.015	0.013	0.063	0.020	0.017
	CM_H	0.600	0.718	0.876	0.169	0.255	0.359	0.116	0.201	0.284
	CM_{OPG}	0.083	0.093	0.124	0.074	0.093	0.117	0.098	0.122	0.149
τ_3	COX_{NCR}	0.353	0.252	0.255	0.456	0.305	0.336	0.641	0.435	0.467
	COX_{CR}	0.041	0.029	0.022	0.159	0.125	0.155	0.383	0.279	0.326
	CM_H	0.501	0.713	0.893	0.180	0.444	0.674	0.223	0.223	0.570
	CM_{OPG}	0.131	0.173	0.240	0.281	0.400	0.505	0.492	0.664	0.755

Quadro 4.2.2 - Potência dos testes quando o parâmetro γ é função do vector de regressores Z_2.

		$N_1 = 100$			$N_2 = 500$			$N_3 = 1000$		
		π_1	π_2	π_3	π_1	π_2	π_3	π_1	π_2	π_3
τ_1	COX_{NCR}	0.204	0.073	0.060	0.104	0.022	0.015	0.082	0.013	0.006
	COX_{CR}	0.041	0.019	0.021	0.047	0.010	0.008	0.047	0.008	0.004
	CM_H	0.693	0.758	0.870	0.254	0.241	0.284	0.145	0.136	0.139
	CM_{OPG}	0.083	0.085	0.104	0.058	0.058	0.060	0.051	0.054	0.054
τ_2	COX_{NCR}	0.259	0.126	0.113	0.220	0.071	0.069	0.261	0.076	0.070
	COX_{CR}	0.042	0.020	0.015	0.044	0.016	0.015	0.079	0.027	0.026
	CM_H	0.561	0.707	0.874	0.136	0.235	0.380	0.096	0.203	0.323
	CM_{OPG}	0.092	0.102	0.139	0.083	0.104	0.141	0.119	0.149	0.194
τ_3	COX_{NCR}	0.390	0.316	0.356	0.559	0.447	0.522	0.776	0.636	0.688
	COX_{CR}	0.047	0.037	0.036	0.258	0.236	0.328	0.573	0.492	0.583
	CM_H	0.454	0.701	0.906	0.210	0.574	0.820	0.340	0.786	0.943
	CM_{OPG}	0.175	0.260	0.374	0.439	0.645	0.774	0.708	0.903	0.958

Com base nos resultados apresentados tirámos as seguintes conclusões:

- De um modo geral, os resultados são bastante fracos. Para valores de $\tau = 0.1$, existem situações em que a percentagem de rejeição é de apenas 0.5%. É importante notar que para a versão H do teste CM, os resultados são bastante enganadores, uma vez que as elevadas percentagens de rejeição devem-se, mais uma vez, à existência de proporções elevadas de estatísticas de teste negativas. Neste mesmo teste, acontece algo que não é razoável que aconteça, ou seja, a frequência de rejeição diminui à medida que a dimensão da amostra aumenta, no entanto, esta diminuição acontece porque com o aumento da dimensão da amostra diminui a percentagem de estatísticas de teste negativas;

- À medida que o valor de τ aumenta, as frequências de rejeição também aumentam, o que significa que este teste é sensível ao afastamento da hipótese nula;

- Contrariamente ao que se verificou na secção anterior, o teste ao modelo de Cox apresentou resultados bastante fracos;

- A existência de observações censuradas faz aumentar a potência dos testes;

- A versão OPG do teste CM foi a que apresentou melhores resultados;

- Os resultados obtidos sugerem que o teste do modelo de Cox tem pouca potência (pelo menos, para algumas alternativas).

5 Conclusões

Neste trabalho fizemos uma breve apresentação sobre modelos econométricos de duração, com especial ênfase no modelo de *Hazard* Proporcional. Na nossa breve apresentação teórica sobre modelos econométricos de duração não fomos exaustivos sobre o tema. No entanto, tentámos seleccionar o conjunto de resultados essenciais quer à elaboração de algum trabalho aplicado com base neste tipo de técnicas quer à auto-suficiência do presente texto. Relativamente ao modelo de *Hazard* Proporcional, apresentámos alguns testes de especificação às hipóteses deste modelo, os quais organizámos em testes formais e testes não formais. Os testes formais foram ainda divididos em testes específicos (entenda-se testes a uma hipótese em particular) e testes gerais.

Numa segunda parte do trabalho, através de simulação de Monte Carlo comparámos a performance em pequenas amostras dos testes de especificação gerais no contexto da estimação por Máxima Verosimilhança e no contexto do modelo de Cox. Deste exercício concluímos que em termos de dimensão ambos os testes apresentam bons resultados, em particular, a versão OPG do teste CM e a versão corrigida para pequenas amostras do teste ao modelo de Cox. Em termos de potência, apenas analisámos a potência do teste contra uma forma particular de não proporcionalidade da *Hazard*. Neste caso, as conclusões foram bastante diferentes das conclusões relativas à dimensão dos testes, ou seja, todos os testes apresentaram resultados fracos, em particular os testes ao modelo de Cox. Um resultado algo inesperado deste exercício foi a má prestação da versão H do teste CM em comparação com a versão OPG do mesmo teste.

Referências Bibliográficas

1. Casella, G. e Berger, R. (2002), *Statistical Inference*, 2^{nd} Edition, Duxbury Press.

2. Chesher, A. (1984), Testing for Neglected Heterogeneity, *Econometrica*, 52(4), pp. 865-872.

3. Chesher, A., Lancaster, T. e Irish, M. (1985), On Detecting the Failure of Distributional Assumptions, *Annales de L'INSEE*, 59/60, pp. 7-44.

4. Cox, D. (1972), Regression Models and Life Tables (with discussion), *Journal of the Royal Statistical Society*, B 34, pp. 187-220.

5. Cox, D. e Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall.

6. Elbers, C. e Ridder, G. (1982), True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model, *The Review of Economic Studies*, vol. 49(3), pp 403-09.

7. Godfrey, L. G. (1988), *Misspecification Tests in Econometrics. The Lagrange Multiplier Principle and other approaches*, Econometric Society Monographs, 16, Cambridge University Press.

8. Heckman, J. e Singer, B. (1984), Econometric Duration Analysis, *Journal of Econometrics*, Annals 1984-1, pp. 63-132.

9. Horowitz, J.L. e Neumann, G.R. (1989), Specification Testing in Censored Regression Models: Parametric and Semiparametric Methods, *Journal of Applied Econometrics*, Vol.4, pp. S61-S86.

10. Horowitz, J.L. e Neumann, G.R. (1992), A Generalized Moments Specification Test of the Proportional Hazard, *Journal of the American Statistical Association*, 87(417), pp. 234-240.

11. Jaggia, S. (1991), Tests of Moments Restrictions in Parametric Duration Models, *Economics Letters*, 37, pp. 35-38.

12. Jaggia, S. (1997), Alternative forms of the Score Test for Heterogeneity, *The Review of Economics and Statistics*, May 1997, pp. 340-343.

13. Jaggia, S. e Trivedi, P. (1994), Joint and Separate Score Tests for State Dependence and Unobserved Heterogeneity, *Journal of Econometrics*, 60, pp. 273-291.

14. Johnson, N., Kotz, S. e Balakrishnan, N. (1994), *Continuous Univariate Distributions*, Vol. 1, second edition, Wiley Series in Probability and Mathematical Statistics.

15. Kiefer, N. (1988), Economic Duration Data and Hazard Functions, *Journal of Economic Literature*, Vol. XXVI, June, pp. 646-679.

16. Lancaster, T. (1979) Econometric Methods for the Duration of Unemployment, *Econometrica*, 47(4), pp. 939-956.

17. Lancaster, T. (1985), Generalized Residuals and Heterogenous Duration Models: with application to the Weibull Model, *Journal of Econometrics*, 28, pp. 155-169.

18. Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Econometric Society Monographs, 17, Cambridge University Press.

19. Lancaster, T. e Chesher, A. (1984), Residual Analysis of Censored Duration Data, *Discussion Paper Bristol University*, No. 84/159.

20. Lancaster, T. e Chesher, A. (1985a), Residuals, Tests and Plots with a Job Matching Illustration, *Annales de L"INSEE*, No. 59/60, pp. 47-69.

21. Lancaster, T. e Chesher, A. (1985b), Residual Analysis of Censored Duration Data, *Economics Letters*, 18, pp. 35-38.

22. McCall, B. P. (1994), Testing the Proportional Hazards Assumption in the presence of Unmeasured Heterogeneity, *Journal of Applied Econometrics*, Vol.9, pp. 321-334.

23. Murteira, B., Ribeiro, C.S., Silva, J.A. e Pimenta, C. (2002), *Introdução à Estatística*, McGraw Hill.

24. Neumann, G.R. (1997), Search Models and Duration Data, Chapter 7 in *Hanbook of Applied Econometrics: Microeconometrics*, M. H. Peasaran, ed. (Basil Blackwell: Oxford), pp. 300-351.

25. Newey, W. K. (1985), Maximum Likelihood Specification Testing and Conditional Moments Tests, *Econometrica*, 53, pp. 1047-1070.

26. Pagan, A. e Ullah, A. (1999), *Nonparametric Econometrics*, Themes in Modern Econometrics, 6, Cambridge.University Press.

27. Pagan, A. e Vella, F. (1989), Diagnostic Tests for Models based on Individual Data: a Survey, *Journal of Applied Econometrics*, Vol. 4, pp. S29-S59.

28. Ramsey, J.B. (1969), Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis, *Journal of the Royal Statistical Society*, Series B, 31, pp. 350-371.

29. Santos Silva, J. M. C. (2001), A Score Test for Non-Nested Hypotheses with applications to Discrete Data Models, *Journal of Applied Econometrics*, vol. 16, pp. 577-597.

30. Tauchen, G. (1985), Diagnostic Testing and Evaluation of Maximum Likelihood Models, *Journal of Econometrics*, 30, pp. 415-443.

31. Tsiatis, A. (1981), A Large Sample Study of Cox's Regression Model, *The Annals of Statistics*, Vol.9, No. 1, pp. 93-108.

32. Wooldridge, J. M. (1992), Some alternatives to the Box-Cox Regression Model, *International Economic Review*, Vol. 33, No. 4, pp. 935-955.

33. Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT press.

Apêndice 1A - Testes de Wald, LR, LM, $C(\alpha)$ e CM

O apêndice 1A tem como objectivo apresentar de forma sucinta e tão intuitiva quanto possível as 5 formas de teste que foram sendo mencionadas ao longo do texto.

Suponha-se que se tem uma amostra aleatória de dimensão N e θ, um vector de parâmetros desconhecidos de dimensão $r \times 1$, é estimado por algum método $N^{\frac{1}{2}}$ consistente. Suponha-se ainda que se pretende testar a seguinte hipótese:

$$H_0 : R(\theta_0) = 0, \tag{51}$$

onde $R(\cdot)$ é uma função vectorial (linear ou não linear) $m \times 1$ - $m < r$ - do vector de parâmetros θ. Para realizar este teste existem na literatura pelo menos cinco metodologias: testes de Wald, testes de Rácio de Verosimilhanças (LR), testes de Multiplicadores de Lagrange ou *Score* (LM), testes de *Score* condicional ($C(\alpha)$) e testes de Momentos Condicionais(CM). Uma característica importante destas cinco metodologias é o facto de serem assimptoticamente equivalentes[19]. Apesar da equivalência assimptótica, em amostras finitas podem existir algumas diferenças assim como, em determinadas situações, pode ser preferível a utilização de um dos testes em detrimento dos outros quer por ser mais simples quer pelo facto de a hipótese a testar não permitir a utilização de algum(ns) teste(s).

[19] Veja-se Godfrey (1988) para uma exposição detalhada das várias hipóteses que são necessárias verificar para que os testes sejam válidos e também para a demonstração dos mesmos.

1A.1 Teste de Wald

O teste de Wald[20] consiste em avaliar a hipótese que se pretende testar utilizando as estimativas obtidas para θ, i.e., a realização do teste de Wald pressupõe apenas que o modelo em causa seja estimado sem que a hipótese que se pretende testar - H_0 - seja imposta no modelo.

Seja $\widehat{\theta}$ um estimador consistente de θ, $V(\widehat{\theta})$ a matriz de variâncias-covariâncias assimptótica do mesmo estimador e $R^{(1)}(\theta)$ é a matriz Jacobiana de $R(\theta)$, nestas condições, a estatística de Wald é dada pela seguinte expressão:

$$sob\ H_0: \ W \equiv R(\theta)'[R^{(1)}(\theta)V(\widehat{\theta})R^{(1)}(\theta)']^{-1}R(\theta) \overset{a}{\sim} \chi^2_{(m)}. \qquad (52)$$

1A.2 Teste de Rácio de Verosimilhanças

O teste de Rácio de Verosimilhanças[21] (LR) foi inicialmente proposto no contexto de estimação por Máxima Verosimilhança. No entanto, com base na teoria dos estimadores extremos, é possível mostrar que o teste LR é um caso particular dos testes de comparação de funções objectivo. Dada a hipótese que se pretende testar é necessário definir dois modelos: o primeiro modelo incorpora a hipótese nula - modelo restrito - e o segundo incorpora a hipótese alternativa - modelo livre. O teste LR consiste em medir quão mais "verosímil" é o modelo livre em comparação com o modelo restrito. Para efectuar esta comparação, calcula-se a diferença entre as funções objectivo dos dois modelos. Caso a hipótese nula seja aceitável, as duas

[20] Wald, A. (1943) "Tests of Statistical hypotheses concerning several parameters when the number of observations is large", Transactions of the American Mathematical Society, 54:426-482.

[21] Wilks, S. S. (1938), "The large sample distribution of the likelihood ratio for testing composite hypotheses.", Annals of Mathematical Statistics, 9:60-62.

funções objectivo não deverão ser muito diferentes. O teste LR tem a seguinte forma:

$$sob\ \ H_0: \ \ LR \equiv 2N(\widehat{Q}_{restrito} - \widetilde{Q}_{livre}) \overset{a}{\sim} \chi^2_{(m)}. \tag{53}$$

No caso de estimação por Máxima Verosimilhança, $\widehat{Q}_{restrito}$ e $\widetilde{Q}_{livre}$ devem ser substituídos por $\frac{1}{N}l(\widehat{\theta})$ e $\frac{1}{N}l(\widetilde{\theta})$ respectivamente e a expressão (53) para o caso do estimador ser de Máxima Verosimilhança é:

$$sob\ \ H_0: \ \ LR \equiv 2(l(\widehat{\theta}_{restrito}) - l(\widetilde{\theta}_{livre})) \overset{a}{\sim} \chi^2_{(m)}. \tag{54}$$

1A.3 Teste de Multiplicadores de Lagrange

Tal como já foi referido anteriormente, os testes LM[22] são um caso particular dos testes CM, no entanto, ambos testes serão apresentados por forma a mostrar a sua ligação. Suponha-se que o modelo $f(\cdot)$ resulta do modelo $g(\cdot)$ quando a hipótese nula é imposta e que os parâmetros do modelo $f(\cdot)$ são estimados por Máxima Verosimilhança. Nestas condições, o teste LM consiste em testar a nulidade dos *scores* (derivadas parciais da função logaritmo da verosimilhança) do modelo $g(\cdot)$ avaliados nas estimativas dos parâmetros do modelo $f(\cdot)$ e nas restrições que resultam da hipótese nula. Como os *scores* correspondentes aos parâmetros estimados são identicamente 0, devido às condições de primeira ordem do estimador de Máxima Verosimilhança, apenas se testa a nulidade dos *scores* referentes aos parâmetros

[22] Rao, C.R. (1948), "Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation.", Proceedings of the Cambridge Philosophical Society, 44:50 - 57.

sobre os quais incide a hipótese nula. O teste LM é da seguinte forma:

$$sob\ H_0 : LM = \widehat{s}_r'[V(\widehat{\theta}_r)]^{-1}\widehat{s}_r \overset{a}{\sim} \chi^2_{(m)}, \tag{55}$$

onde $\widehat{s}_r$ representa o vector $m \times 1$ dos scores referentes a H_0.

1A.4 Testes de Momentos Condicionais

Em 1985, em trabalhos separados, Tauchen[23] e Newey[24] desenvolveram uma metodologia de teste que permitiu generalizar os testes LM assim como os testes da matriz de informação de White. A ideia fundamental deste tipo de testes é que qualquer hipótese que se queira testar pode ser expressa sob a forma de momento/valor esperado, i.e., dada a hipótese a testar existe uma variável aleatória cujos momentos dependem da hipótese nula. Desta forma, o valor teórico desses momentos poderá ser calculado e comparado com o valor dos momentos amostrais correspondentes. Seja $R(\theta_0)$ o conjunto de m momentos que se pretende testar e seja θ_0^* um vector de parâmetros e/ou estimativas. Nestas condições, o teste CM é da seguinte forma:

$$sob\ H_0 : CM = R(\theta_0^*)'[V(R(\theta_0^*))]^{-1}R(\theta_0^*) \overset{a}{\sim} \chi^2_{(m)}. \tag{56}$$

[23]Tauchen, G. (1985), "Diagnostic testing and evaluation of maximum likelihood models.", Journal of Econometrics 30, 415-443.

[24]Newey, W.K. (1985), "Maximum likelihood specification testing and conditional moment tests.", Econometrica 53, 1047-1070.

1A.5 Teste $C(\alpha)$ ou de *Score* Condicional

Os testes $C(\alpha)$ ou de score condicional são certamente menos conhecidos do que qualquer um dos testes apresentados até agora. No entanto, esta metodologia de teste é assimptoticamente equivalente às outras quatro. Esta metodologia foi proposta por Neyman (1959)[25] e recuperada para o contexto de modelos de duração por Jaggia e Trivedi (1994)[26] e, apesar de serem pouco conhecidos, este tipo de testes têm como caso particular os testes LM e os testes CM. A especificação de modelos econométricos, por norma, implica que sejam feitas algumas hipóteses *a priori* que, também por norma, não são independentes entre si. Esta situação faz com que quando se testa uma das hipóteses isoladamente se esteja a supor a validade das outras. Caso não se verifique a validade de uma das hipóteses não testadas pode-se estar a rejeitar uma hipótese nula sem que seja falsa devido a outro erro de especificação. O contributo dos testes $C(\alpha)$ é permitirem testar hipóteses individualmente sem que seja necessário supor a validade das outras, ou seja, os testes $C(\alpha)$ são testes marginais para as hipóteses não testadas.

Sejam $\widetilde{\theta}$ um estimador $N^{\frac{1}{2}}$ consistente para o vector θ, $s(\widetilde{\theta})$ o score avaliado no estimador $\widetilde{\theta}$ - $E[s(\widetilde{\theta})] = 0$ - e $R[\theta_0|s(\widetilde{\theta})]$ a hipótese que se pretende testar mas condicional no valor verificado para $s(\widetilde{\theta})$. Nestas condições, o teste $C(\alpha)$ é da seguinte forma:

$$sob \;\; H_0 : C(\alpha) = R[\theta_0^*|s(\widetilde{\theta})]'[V(R[\theta_0^*|s(\widetilde{\theta})])]^{-1}R[\theta_0^*|s(\widetilde{\theta})] \overset{a}{\sim} \chi^2_{(m)}. \qquad (57)$$

[25]Neyman, J. (1959), "Optimal asymptotic tests of composite statistical hypothesis", U.Grenander ed., Probability and Statistics (Wiley, New York, NY) 213-234.

[26]Jaggia, S. e Trivedi, P. K. (1994), "Joint and separate score tests for state dependence an unobserved heterogeneity.", Journal of Econometrics, 60, 273-291.

Quando $\widetilde{\theta}$ é o estimador de Máxima Verosimilhança, o teste $C(\alpha)$ é idêntico ao teste CM.

Apêndice 1B - Estimadores da matriz de variâncias covariâncias.

No apêndice 1A foram apresentadas cinco formas de teste de uma forma bastante geral sem que tenha sido apresentada a forma de estimação de alguns elementos essenciais na construção de quase todos os testes, em particular, não foi apresentado nenhum estimador para a matriz de variâncias covariâncias de $\widehat{\theta}_{MV}$ - $V[\widehat{\theta}_{MV}]$. Neste apêndice, serão apresentados três estimadores para $V[\widehat{\theta}_{MV}]$.

Seja $\widehat{\theta}_{MV}$ o estimador de máxima verosimilhança de θ, $S(\theta)$ o vector de *scores* de θ, $H(\theta)$ a matriz hessiana da função logaritmo de verosimilhança (ou a matriz de primeiras derivadas do vector $S(\theta)$), $A = -E[H(\theta)]$ e $B = E[S(\theta)S(\theta)']$, sob determinadas condições de regularidade (as mesmas referidas no apêndice anterior):

$$\sqrt{N}(\widehat{\theta}_{MV} - \theta) \overset{a}{\sim} N(0, A^{-1}BA^{-1}). \tag{58}$$

Dada a igualdade da matriz de informação condicional[27] a matriz de variâncias covariâncias do estimador MV simplifica-se bastante, ou seja,

$$\sqrt{N}(\widehat{\theta}_{MV} - \theta) \overset{a}{\sim} N(0, A^{-1}). \tag{59}$$

Desta forma, estimar $V[\widehat{\theta}_{MV}]$ implica estimar A ou B.

O primeiro método consiste em encontrar a matriz A ou a matriz B e substituir

[27] $-E[H(\theta)|X] = E[S(\theta)S(\theta)'|X] \Rightarrow A = B$ (pela regra do valor esperado iterado),

θ por $\widehat{\theta}_{MV}$. Neste caso,

$$\widehat{V[\widehat{\theta}_{MV}]} = [\sum_{i=1}^{N} A(\widehat{\theta}_{MV}, X)]^{-1}. \tag{60}$$

Apesar de este método ser o que tem melhores propriedades, por norma, não é possível obter a matriz A ou a matriz B quando existe censura uma vez que a função de distribuição conjunta do tempo - T - e da censura - C - não é conhecida. Este facto faz com que na prática não seja possível utilizar este estimador. Tirando partido da igualdade da matriz de informação surgem dois métodos, eventualmente com piores propriedades do que o primeiro, mas cuja aplicação é relativamente simples. Desta forma, é possível estimar consistentemente a matriz $V[\widehat{\theta}_{MV}]$ quer avaliando a matriz hessiana na estimativa de θ

$$\widehat{V[\widehat{\theta}_{MV}]} = -[\sum_{i=1}^{N} H(\widehat{\theta}_{MV}, X)]^{-1}, \tag{61}$$

quer calculando o produto externo do vector $S(\theta)$ avaliado na mesma estimativa

$$\widehat{V[\widehat{\theta}_{MV}]} = [\sum_{i=1}^{N} S(\widehat{\theta}_{MV}, X)S(\widehat{\theta}_{MV}, X)']^{-1}. \tag{62}$$

Embora ambos os estimadores sejam consistentes para a matriz $V[\widehat{\theta}_{MV}]$, em amostras finitas, a escolha do estimador poderá influenciar as propriedades dos testes. Por razões óbvias, a única metodogia de teste que não tem as suas propriedades dependentes desta escolha é a metodologia LR. Todas as outras dependem do estimador da matriz de variâncias-covariâncias utilizado.

Apêndice 1C - testes de momentos condicionais no contexto do modelo de Cox

Devido ao facto de o modelo de Cox não ser estimado por Máxima Verosimilhança, os testes de momentos condicionais apresentados anteriormente não são válidos neste contexto. No entanto, Horowitz e Neumann (1992) desenvolveram a metodologia de teste equivalente aos testes de momentos condicionais clássicos (Tauchen e Newey) no contexto do modelo de Cox.

Sejam (T, X, D) uma amostra aleatória de dimensão N e $\widehat{\beta}_{VP}$ o estimador de verosimilhança parcial do vector β de parâmetros do seguinte modelo:

$$\lambda(t|x, v, \beta) = \lambda_0(t) f(x, v, \beta). \tag{63}$$

Tsiatis (1981) demonstrou a consistência e a normalidade assimptótica do estimador $\widehat{\beta}_{VP}$,

$$\sqrt{N}(\widehat{\beta}_{VP} - \beta) \overset{a}{\sim} N(0; V_{\beta\beta}). \tag{64}$$

Onde $V_{\beta\beta}$ representa a matriz de variâncias-covariâncias de $\widehat{\beta}_{VP}$. Esta matriz obtém-se através do negativo da inversa da matriz de Informação, i.e., $V_{\beta\beta} = -[I(\beta)]^1 = -\left[\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'}\right]^{-1}$.

Uma vez que a estimação de β pela metodologia de Cox não requer a especificação da função baseline *hazard* - $\lambda_0(t)$ - ou, equivalentemente, a função baseline *hazard* integrada - $\Lambda_0(t)$ - mas, a realização de testes de momentos condicionais implica que se especifique o modelo como um todo, é necessário estimar $\lambda_0(t)$ ou $\Lambda_0(t)$. No mesmo artigo de 1981, Tsiatis demonstrou a consistência e normalidade assimptótica de um estimador de $\Lambda_0(t)$. Este resultado torna natural a escolha entre estimar $\lambda_0(t)$

ou $\Lambda_0(t)$. O estimador de $\Lambda_0\,(t)$ é:

$$\widehat{\Lambda}_0(t) = \sum_{T_i \leq t}\left[\frac{(1-d_i)}{\sum_{j\in R(T_i)} f(x_j, \widehat{\beta}_{VP})}\right], \tag{65}$$

$$\sqrt{N}\left(\widehat{\Lambda}_0(t) - \Lambda_0(t)\right) \overset{a}{=} N\left(0, V_1(t)\right). \tag{66}$$

Sendo $R(T_i)$ o conjunto de observações que no momento t ainda não terminaram ou que não foram censuradas até esse mesmo momento e $V_1(t)$ a variância assimptótica do estimador $\widehat{\Lambda}_0(t)$.

Ao contrário dos testes de momentos condicionais propostos por Tauchen (1985) e Newey (1985), o teste de momentos condicionais desenvolvido por Horowitz e Neumann (1992) não se aplica directamente a todos os tipos de momentos condicionais. Neste caso, os testes foram pensados para a realização de testes sobre os resíduos generalizados. Sejam $e\,(T, X, D, \beta) = \Lambda_0\,(t)\exp\left(X'\beta\right)$ o erro generalizado, $\widehat{e}\left(T, X, D, \widehat{\beta}_{VP}\right) = \widehat{\Lambda}_0\,(t)\exp\left(X'\widehat{\beta}_{VP}\right)$ o resíduo generalizado e $W\left[\widehat{e}\left(T, X, C, \widehat{\beta}_{VP}\right)\right]$ alguma função vectorial $l \times 1$ do resíduo generalizado, tal que, quando o modelo está correctamente especificado, se verifica o seguinte:

$$E\left\{W\left[\widehat{e}\left(T, X, D, \widehat{\beta}_{VP}\right)\right]\right\} = 0. \tag{67}$$

Se o resultado anterior se verificar, então,

$$\Omega_n = n^{-\frac{1}{2}}\sum_{i=1}^{n} W_i\left[\widehat{e}_i\left(T_i, X_i, D_i, \widehat{\beta}_{VP}\right)\right] \overset{a}{=} N\left(0; V_\Omega\right), \tag{68}$$

logo

$$\Omega_n' V_\Omega^{-1}\Omega_n \overset{a}{=} \chi^2_{(l)}. \tag{69}$$

V_Ω é a matriz de variâncias-covariâncias assimptótica de $W\left[\widehat{e}\left(T, X, C, \widehat{\beta}_{VP}\right)\right]$. Apesar do resultado da equação (69) ser válido assimptoticamente, em pequenas amostras poderá ter alguns problemas, pelo que os autores sugeriram que se utilizasse uma forma alternativa de Ω_n, com uma correcção para amostras finitas,

$$(\Omega_n - b_n)' V_\Omega^{-1} (\Omega_n - b_n) \overset{a}{\sim} \chi^2_{(l)}. \tag{70}$$

No que se segue iremos apresentar a forma de cálculo de V_Ω e de b_n.

Obtenção da matriz V_Ω

A obtenção da matriz V_Ω não é única. Os autores do artigo, por exemplo, apresentam uma forma de cálculo desta matriz no artigo em causa e numa aplicação em Gauss, desenvolvida pelos próprios, utilizam outra forma de cálculo. No que se segue, apresentamos a forma de cálculo de V_Ω utilizada na aplicação já referida, uma vez que todos os resultados de simulação de Monte Carlo apresentados neste trabalho foram obtidos utilizando esta mesma aplicação.

A matriz V_Ω resulta do produto interno de três elementos,

$$V_\Omega = E\left[(W + C + A - E[W + C + A])'(W + C + A - E[W + C + A])\right]. \tag{71}$$

O primeiro elemento, W, não apresenta quaisquer dificuldades de cálculo uma vez que é o vector das condições de momentos que estão a ser testadas. O segundo e terceiro elementos, C e A, são algo mais complexos em termos de cálculo. Antes de apresentar as fórmulas em concreto, vamos definir primeiro alguns elementos necessários ao cálculo, assim como estimadores possíveis para os mesmos. Na exposição que se segue, todos os dados estão ordenados por ordem crescente em T_i, ou seja, $T_{i-1} < T_i$.

Sejam

$$P(T,X,C) \text{ a função cumulativa de probabilidade de } (T,X,C), \tag{72}$$

$$\widehat{P}(T_i,X_i,D_i)=\frac{i}{n}, \tag{73}$$

$$W_e[e(t,x,d,\beta)]=\frac{\partial W[e(t,x,d,\beta)]}{\partial e}, \tag{74}$$

$$\widehat{W}_e\left[\widehat{e}\left(T_i,X_i,D_i,\widehat{\beta}_{VP}\right)\right]=\left.\frac{\partial W[e(t,x,d,\beta)]}{\partial e}\right|_{\widehat{e}\left(T_i,X_i,D_i,\widehat{\beta}_{VP}\right)}, \tag{75}$$

$$g(z)=\exp(z'\beta), \tag{76}$$

$$\widehat{g}(z)=\exp\left(z'\widehat{\beta}_{VP}\right), \tag{77}$$

$$D(t)=E[g(X)I(T\geq t)], \tag{78}$$

$$\widehat{D}(t)=\frac{\sum_{i=1}^{n}[\widehat{g}(X_i)I(T_i\geq t)]}{n}, \tag{79}$$

$$D_x(t)=E[Xg(X)I(T\geq t)], \tag{80}$$

$$\widehat{D}_x(t)=\frac{\sum_{i=1}^{n}[X_i\widehat{g}(X_i)I(T_i\geq t)]}{n}, \tag{81}$$

$$Q(t)=E[(1-D)I(T\geq t)]=P(T\geq t,D=0), \tag{82}$$

$$\widehat{Q}(t)=\frac{\sum_{i=1}^{n}[(1-D_i)I(T_i\geq t)]}{n}, \tag{83}$$

$$\phi(t,x)=x\Lambda_0(t)+\int_0^t\frac{D_x(t')}{[D(t')]^2}dQ(t'), \tag{84}$$

$$\widehat{\phi}(t,x)=x\widehat{\Lambda}_0(t)+\sum_{T_i\leq t}\frac{\widehat{D}_x(T_i)}{\left[\widehat{D}(T_i)\right]^2}\Delta\widehat{Q}(T_i), \tag{85}$$

$$R_1(t)=\int_0^t\frac{1}{[D(t')]^2}dQ(t'), \tag{86}$$

$$\widehat{R}_1(t)=\sum_{T_i\leq t}\frac{1}{\left[\widehat{D}(T_i)\right]^2}\Delta\widehat{Q}(T_i), \tag{87}$$

$$R_2(t)=\int_0^t\frac{1}{[D(t')]^2}dD(t'), \tag{88}$$

$$\widehat{R}_2(t) = \sum_{T_i \leq t} \frac{1}{\left[\widehat{D}(T_i)\right]^2} \Delta \widehat{D}(T_i), \tag{89}$$

$$\begin{aligned} S_1(t) &= g(x)\{E[I(T \leq t)\, g(x)\, W_e(t,x,d)\, R_1(t)] - \\ &\quad (1-D)\, E[I(T \leq t)\, g(x)\, W_e(t,x,d)\, R_2(t)]\}, \end{aligned} \tag{90}$$

$$\begin{aligned} \widehat{S}_1(t) &= \widehat{g}(X_i) \left\{ \frac{\sum_{T_i \leq t} \left[\widehat{g}(X_i)\, \widehat{W}_e(T_i, X_i, D_i)\, \widehat{R}_1(t)\right]}{n} - \right. \\ &\quad \left. (1-D_i) \frac{\sum_{T_i \leq t} \left[\widehat{g}(X_i)\, \widehat{W}_e(T_i, X_i, D_i)\, \widehat{R}_2(t)\right]}{n} \right\}, \end{aligned} \tag{91}$$

$$S_2(t) = E\left[I(T = t-1) \int_t^{t^{\max}} W_e(t,x,d)\, g(x)\, dt \right] \{g(x)\, R_1(t) - (1-d)\, R_2(t)\}, \tag{92}$$

$$\widehat{S}_2(t) = \frac{\sum_{T_i \geq t}^{T^{\max}} I(T_i = t-1)\, \widehat{W}_e(T_i, X_i, D_i)\, g(X_i)}{n} \left\{ \widehat{g}(X_i)\, \widehat{R}_1(t) - (1-D_i)\, \widehat{R}_2(t) \right\}, \tag{93}$$

$$S_3(t) = (1-d) \left\{ \frac{E[W_e(t,x,d)\, g(x)]}{E[g(x)]} - E\left[\int_0^t W_e(t,x,d)\, g(x) \frac{1}{D(t)} dt \right] \right\}, \tag{94}$$

$$\widehat{S}_3(t) = (1-D_i) \left\{ \frac{\sum_{i=1}^n \widehat{W}_e(T_i, X_i, D_i)\, \widehat{g}(X_i)}{\sum_{i=1}^n \widehat{g}(X_i)} - \frac{\sum_{T_i \leq t} \widehat{W}_e(T_i, X_i, D_i)\, \widehat{g}(X_i) \frac{1}{\widehat{D}(T_i)}}{n} \right\}, \tag{95}$$

$$V_1(t_1, x_1; t_2, x_2) = [g(x_1)\, g(x_2)] \left\{ -R_1(t^{"}) + \phi(t_1, x_1)' V_{\beta\beta} \phi(t_2, x_2) \right\}, \tag{96}$$

$$\widehat{V}_1(t_1, x_1; t_2, x_2) = [\widehat{g}(x_1)\, \widehat{g}(x_2)] \left\{ -\widehat{R}_1(t^{"}) + \widehat{\phi}(t_1, x_1)' \widehat{V}_{\beta\beta} \widehat{\phi}(t_2, x_2) \right\}, \tag{97}$$

$$\begin{aligned} SMC(t) &= (1-d) \left\{ x - \frac{D_x(t)}{E[g(x)]} - \int_0^t \frac{1}{D(t)} dD_x(0) + \int_0^t \frac{D_x(t)}{[D(t')]^2} dD(t) \right\} - \\ &\quad g(x) \int_0^t \frac{D_x(t)}{[D(t')]^2} dQ(t), \end{aligned} \tag{98}$$

$$\begin{aligned} \widehat{SMC}(t) &= (1-D_i) \left\{ X_i - \frac{\widehat{D}_x(T_i)}{\sum_{i=1}^n \frac{\widehat{g}(X_i)}{n}} - \sum_{T_i \leq t} \frac{1}{\widehat{D}(T_i)} \Delta \widehat{D}_x(0) + \sum_{T_i \leq t} \frac{\widehat{D}_x(T_i)}{\left[\widehat{D}(T_i)\right]^2} \Delta \widehat{D}(T_i) \right\} - \\ &\quad \widehat{g}(X_i) \sum_{T_i \leq t} \frac{\widehat{D}_x(T_i)}{\left[\widehat{D}(T_i)\right]^2} \Delta \widehat{Q}(T_i). \end{aligned} \tag{99}$$

Com base nas funções definidas anteriormente, os elementos C e A obtêm-se da seguinte forma:

$$C = \{E\left[W_e(t,x,d)\,\phi(t,x)'\,V_{\beta\beta}g(x)\right] SMC(t)\}', \tag{100}$$

$$\widehat{C} = \left\{\frac{\sum_{i=1}^{n}\widehat{W}_e(T_i,X_i,D_i)\,\widehat{\phi}(T_i,X_i)'\,\widehat{V}_{\beta\beta}\widehat{g}(X_i)}{n}\widehat{SMC}(t)\right\}', \tag{101}$$

$$A = S_1(t) + S_2(t) + S_3(t), \tag{102}$$

$$\widehat{A} = \widehat{S}_1(t) + \widehat{S}_2(t) + \widehat{S}_3(t). \tag{103}$$

Uma vez estimados os elementos W, C e A, para estimar a matriz V_Ω basta substituir cada um dos elementos pelos estimadores correspondentes,

$$\widehat{V}_\Omega = \frac{\sum_{i=1}^{n}\left(\widehat{W}_i + \widehat{C}_i + \widehat{A}_i - \frac{\sum_{i=1}^{n}\widehat{W}_i+\widehat{C}_i+\widehat{A}_i}{n}\right)'\left(\widehat{W}_i + \widehat{C}_i + \widehat{A}_i - \frac{\sum_{i=1}^{n}\widehat{W}_i+\widehat{C}_i+\widehat{A}_i}{n}\right)}{n} \tag{104}$$

5.1 Obtenção do elemento de correcção b_n

O elemento de correcção b_n e o seu estimador são dados pelas seguintes expressões:

$$b_n = \frac{1}{2}n^{-\frac{1}{2}}E\left[W_e(t,x,d)\,V_1(t,x;t,x)\right], \tag{105}$$

$$\widehat{b}_n = \frac{1}{2}n^{-\frac{1}{2}}\frac{\sum_{i=1}^{n}\widehat{W}_e(T_i,X_i,D_i)\,\widehat{V}_1(T_i,X_i;T_i,X_i)}{n}. \tag{106}$$

Apêndice 2A - Condições de momentos

O apêndice 2A tem como objectivo apresentar as condições de momentos de cada um dos testes apresentados ao longo do texto. Utilizando a notação anterior, as condições de momentos são as seguintes:

Condição 1 - teste à proporcionalidade da *hazard* e teste Reset

$$H_0 : \beta_j = 0, \tag{107}$$

$$C_1 : \left.\left\{(1-d)\frac{\frac{\partial f(x,v,\beta)}{\partial \beta_j}}{f(x,v,\beta)} - \Lambda_0(t)\frac{\partial f(x,v,\beta)}{\partial \beta_j}\right\}\right|_{\beta_j=0}. \tag{108}$$

Condição 2 - teste à forma funcional de $f(x,v,\beta)$

$$H_0 : f(x,v,\beta) = \exp(x'\beta), \tag{109}$$

$$C_2 : \left\{-\frac{1}{2}(x'\beta)^2\left[(1-d) - \Lambda_0(t)\exp(x'\beta)\right]\right\} \Longleftrightarrow (x'\beta)^2\left[(1-d) - \Lambda_0(t)\exp(x'\beta)\right]. \tag{110}$$

Condição 3 - teste à forma funcional de $\lambda_0(t)$, alternativas não encaixadas

$$H_0 : \alpha = 0, \tag{111}$$

$$C_{3,1} : \left[\frac{\lambda_2(t|x,v,\theta)}{\lambda_1(t|x,v,\theta)}\right]^{1-d}\exp\{-\Lambda_2(t|x,v,\theta) + \Lambda_1(t|x,v,\theta)\} - 1. \tag{112}$$

$$H_0 : \alpha = 1, \tag{113}$$

$$C_{3,2} : \left[\frac{\lambda_1(t|x,v,\theta)}{\lambda_2(t|x,v,\theta)}\right]^{1-d}\exp\{-\Lambda_1(t|x,v,\theta) + \Lambda_2(t|x,v,\theta)\} - 1. \tag{114}$$

Condição 4 - teste à existência de heterogeneidade negligenciada

$$H_0 : \sigma_v^2 = 0, \tag{115}$$

$$C_4 : \quad \left\{-\frac{1}{2}\left[2\,(1-d)\,\Lambda\,(t|x,\theta) - \Lambda\,(t|x,\theta)^2\right]\right\} \Longleftrightarrow \left[2\,(1-d)\,\Lambda\,(t|x,\theta) - \Lambda\,(t|x,\theta)^2\right]. \tag{116}$$

Apêndice 2B - Erros e Resíduos

Ao longo deste trabalho constatámos que apesar de os conceitos de erros apresentados na secção 2.7 serem bastante utilizados e referidos, não encontrámos qualquer demonstração das suas propriedades. Sem qualquer pretensão de sermos os primeiros a demonstrar as já referidas propriedades, optámos por deixar um registo das mesmas assim como das condições de momentos que lhe estão associadas.

Para demonstrar que, na ausência de censura, e tem distribuição exponencial unitária basta reparar que a função sobrevivência de uma variável aleatória Y com distribuição exponencial unitária é $\exp\{-y\}$, $P(T > t) \Longleftrightarrow P[\Lambda(T) > \Lambda(t)] = \exp\{-\Lambda(t)\}$ e que ter y ou $\Lambda(t)$ é indiferente. Logo, $\Lambda(T) \sim Exp(1)$. Utilizando o facto de $E\left[\Lambda(T)^j\right] = j!$[28] é possível criar tantas condições de momentos quantas sejam necessárias. O problema com e é o facto de em dados económicos ser muito raro não existir censura e, quando tal acontece, a distribuição de e deixa de ser $Exp(1)$. Apesar deste problema, é possível construir uma condição de momentos a partir de e tirando partido da função de distribuição de $F(e) = 1 - S(e)$. Um resultado bastante importante em estatística é que qualquer função de distribuição tem como função de distribuição a distribuição Uniforme no intervalo $[0;1]$, i.e., $F[F(Y)] = P[F(Y) \leq F(y)] = y$[29]. Com base neste resultado e na hipótese de censura aleatória e independente, é possível calcular o valor esperado de $F[e]$ ou equivalentemente, de $F[\Lambda(T)]$.

[28] $Gamma[z] = \int_0^{\infty} v^{z-1}\exp\{-z\}dz = (z-1)!$ quando $z \in \mathbb{N}$, substituindo $z-1$ por j obtém-se exactamente $E[\Lambda(T)^j] = \int_0^{\infty} \Lambda(T)^j \exp\{-\Lambda(T)\}dz = Gamma[j+1] = j!$

[29] Para uma demonstração deste resultado veja-se Murteira, Ribeiro, Silva e Pimenta (2002).

Quando não existe censura,

$$E\{F[\Lambda(T)]\} = \int_0^1 F[\Lambda(T)]\,dF[\Lambda(T)] = \frac{1}{2}. \tag{117}$$

No caso em que existe censura,

$$\begin{aligned} E\{F[\Lambda(T)]\,|F[\Lambda(T)] \leq F[\Lambda(C)]\} &= \int_0^{F[\Lambda(C)]} F[\Lambda(T)]\frac{1}{F[\Lambda(C)]}dF[\Lambda(T)] \qquad (118)\\ &= \frac{F[\Lambda(C)]}{2}. \end{aligned}$$

A condição de momentos que resulta das equações (117) e (118) é,

$$\frac{2-D}{2}S[\Lambda(T^*)] - \frac{1}{2} = (2-D)\,S[\Lambda(T^*)] - 1. \tag{119}$$

Um segundo tipo de erros apresentado na secção 2.7 são os erros $\widetilde{e} = e + du$. Estes erros têm como principal característica o facto de mesmo nos casos em que existe censura continuarem a ter distribuição exponencial unitária. A demonstração deste resultado passa por demonstrar que a função geradora de momentos de $\widetilde{e}$ é a função geradora de momentos de uma v.a. exponencial unitária. Se Y for uma v.a. com distribuição exponencial unitária então, a função geradora de momentos de Y é $M_Y(s) = E[\exp(sY)] = (1-s)^{-1}$.

Seja C o ponto de censura,

$$
\begin{aligned}
E\left[\exp\left(s\widetilde{e}\right)\right] &= E_{C,U}\left\{E\left[\exp\left(s\widetilde{e}\right)\right]|C,U\right\} \\
&= E_{C,U}\left\{E\left[\exp\left(s\widetilde{e}\right)\right]|\Lambda\left(T\right)<\Lambda\left(C\right),C,U\right\}P\left[\Lambda\left(T\right)<\Lambda\left(C\right)\right]+ \\
&\quad E_{C,U}\left\{E\left[\exp\left(s\widetilde{e}\right)\right]|\Lambda\left(T\right)\geq\Lambda\left(C\right),C,U\right\}P\left[\Lambda\left(T\right)\geq\Lambda\left(C\right)\right] \\
&= E_U\left\{\int_0^{\Lambda(C)}\frac{\exp\left\{s\Lambda\left(T\right)\right\}\exp\left\{-\Lambda\left(T\right)\right\}}{1-\exp\left\{-\Lambda\left(C\right)\right\}}d\Lambda\left(T\right)\left[1-\exp\left\{-\Lambda\left(C\right)\right\}\right]+\right. \\
&\quad \left.\exp\left\{s\left[\Lambda\left(C\right)+U\right]\right\}\exp\left\{-\Lambda\left(C\right)\right\}|U\right\} \\
&= \int_0^{\infty}\left\{\left[1-s\right]^{-1}\left\{1-\exp\left[\left(s-1\right)\Lambda\left(C\right)\right]\right\}\right. \\
&\quad \left.+\exp\left[\left(s-1\right)\Lambda\left(C\right)\right]\exp\left[sU\right]\exp\left(-U\right)\right\}dU \\
&= \left(1-s\right)^{-1}.
\end{aligned}
$$

Uma vez que $\widetilde{e}\sim Exp(1)$ é possível construir uma infinidade de condições de momentos com base nestes erros.

O terceiro e último tipo de erros apresentado em 2.7 são os erros $e^*=e+d$. Estes erros não têm nenhuma distribuição conhecida. No entanto, é possível determinar os dois primeiros momentos não centrados, logo, também é possível determinar a sua variância. Desta forma, a média de e^* é a seguinte:

$$
\begin{aligned}
E\left[e^*\right] &= E_C\left\{E\left[e^*|C\right]\right\} \\
&= E_C\left\{E\left[e^*\right]|\Lambda\left(T\right)<\Lambda\left(C\right),C\right\}P\left[\Lambda\left(T\right)<\Lambda\left(C\right)|C\right]+ \\
&\quad E_C\left\{E\left[e^*\right]|\Lambda\left(T\right)\geq\Lambda\left(C\right),C\right\}P\left[\Lambda\left(T\right)\geq\Lambda\left(C\right)|C\right] \\
&= \int_0^{\Lambda(C)}\frac{\Lambda\left(T\right)\exp\left(-\Lambda\left(T\right)\right)}{1-\exp\left[-\Lambda\left(C\right)\right]}d\Lambda\left(T\right)\left\{1-\exp\left[-\Lambda\left(C\right)\right]\right\}+ \\
&\quad \left[\Lambda\left(C\right)+1\right]\exp\left[-\Lambda\left(C\right)\right] \\
&= 1.
\end{aligned}
$$

Apesar de este resultado ser bastante interessante, não pode ser utilizado em testes de especificação uma vez que esta condição se verifica sempre, i.e., é uma condição que é transformação linear das condições de primeira ordem dos estimadores[30]. No caso do segundo momento não centrado ou da variância, a propriedade anterior já não se verifica, o que permite que seja utilizado na construção de testes de especificação. O segundo momento de e^* é o seguinte,

$$\begin{aligned}
E\left[(e^*)^2\right] &= E_C\left\{E\left[(e^*)^2 |C\right]\right\} \\
&= E_C\left\{E\left[(e^*)^2\right] |\Lambda(T) < \Lambda(C), C\right\} P\left[\Lambda(T) < \Lambda(C) |C\right] + \\
&\quad E_C\left\{E\left[(e^*)^2\right] |\Lambda(T) \geq \Lambda(C), C\right\} P\left[\Lambda(T) \geq \Lambda(C) |C\right] \\
&= \int_0^{\Lambda(C)} \frac{[\Lambda(T)]^2 \exp(-\Lambda(T))}{1-\exp[-\Lambda(C)]} d\Lambda(T) \{1-\exp[-\Lambda(C)]\} + \\
&\quad [\Lambda(C)+1]^2 \exp[-\Lambda(C)] \\
&= 2 - P[\Lambda(T) \geq \Lambda(C)],
\end{aligned}$$

o que implica que $Var[e^*] = E\left[(e^*)^2\right] - \{E[e^*]\}^2 = 1 - P[\Lambda(T) \geq \Lambda(C)]$. Com base neste resultado, a condição de momentos resultante é:

$$[\Lambda(T^*) + D - 1]^2 - \{1 - P[\Lambda(T) \geq \Lambda(C)]\}. \tag{120}$$

Um aspecto interessante da condição anterior é que, no caso de não existir censura, esta condição de momentos é igual à condição de momentos do teste à existência de heterogeneidade negligenciada - C_3 - e, no caso de existir censura, apenas difere no termo $\{1 - P[\Lambda(T) \geq \Lambda(C)]\}$.

[30]Veja-se Lancaster e Chesher - "Residuals, Tests and Plots with a Job Matching Illustration", (1985), Annales de L'INSEE

Apêndice 3A

Os quadros que se seguem dizem respeito aos valores de M para cada combinação (π, X, γ) utilizados na secção 4.1.

Quadro A.3.1 - Valores utilizados para gerar 20% de observações censuradas - dimensão

$M(\pi_2, X, \gamma)$	X_1	X_2	X_3	X_4
γ_1	1.429	1.414	2.152	0.522
γ_2	1.177	1.027	1.471	0.461
γ_3	1.178	0.955	1.347	0.473

Quadro A.3.2 - Valores utilizados para gerar 50% de observações censuradas - dimensão

$M(\pi_2, X, \gamma)$	X_1	X_2	X_3	X_4
γ_1	0.149	0.175	0.322	0.043
γ_2	0.210	0.263	0.436	0.066
γ_3	0.240	0.306	0.480	0.077

Os quadros que se seguem dizem respeito aos valores de M para cada combinação (π, τ, Z) utilizados na secção 4.2.

Quadro A.3.3 - Valores utilizados para gerar 20% de observações censuradas - potência

$M(\pi_2, \tau, Z)$	τ_1	τ_2	τ_3
Z_1	0.461	0.467	0.485
Z_2	0.461	0.467	0.495

Quadro A.3.4 - Valores utilizados para gerar 50% de observações censuradas - potência

$M(\pi_3, \tau, Z)$	τ_1	τ_2	τ_3
Z_1	0.066	0.064	0.060
Z_2	0.066	0.065	0.061